Contra Chalmers:
On Consciousness
and Conceivability

A Thesis Submitted to the College of
Graduate Studies and Research
in Partial Fulfillment of the Requirements
for the Degree of Master of Arts
in the Department of Philosophy
University of Saskatchewan
Saskatoon

By

Jennifer-Wrae Primmer

PERMISSION TO USE

In presenting this thesis/dissertation in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis/dissertation in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis/dissertation work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis/dissertation or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis/dissertation.

Requests for permission to copy or to make other uses of materials in this thesis/dissertation in whole or part should be addressed to:

Head of the Department of Philosophy
University of Saskatchewan
Saskatoon, Saskatchewan, S7N 5A5
Canada

OR

Dean
College of Graduate Studies and Research
University of Saskatchewan
107 Administration Place
Saskatoon, Saskatchewan, S7N 5A2
Canada

ABSTRACT

This thesis presents and evaluates David Chalmers' argument that the existence of phenomenal conscious experience constitutes a permanent barrier to the reductive aspirations of a purely materialistic neuroscience. My aim is to defend the possibility of a reductive explanation of consciousness, and argue that continued research in neuroscience and neurophysiology can result in a successful materialistic or reductive solution to the hard problem of consciousness. My argument against Chalmers is two-fold. First, I challenge Chalmers' claim that consciousness does not logically supervene on the physical. And second, I argue that his conceivability argument is implausible.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Dr. Sarah Hoffman. This thesis would not have been possible without her exceptional support, guidance, and friendship. I appreciate having had the opportunity to work with Dr. Hoffman, and am grateful for her eye to detail, her compassionate nature, and her dislike of philosophical zombies. I must also thank my committee members Dr. Karl Pfeifer and Dr. Eric Dayton for their valuable insights. I am grateful for Dr. Pfeifer's knowledge of the subject matter and his continual support. I thank him for keeping me on my toes during every discussion. Dr. Dayton's enthusiasm for philosophical discourse and his willingness to answer any question with patience was always helpful. I also thank him for serving as a constant reminder of the type of Professor I aspire to be. I would also like to thank Della Nykyforak for her kindness and support. A special thanks to my fiancé Adam Sopuck, whom I relied on for both academic and emotional support. I also thank him for his patience, and for bringing clarity into my sometimes irrational world.

I am especially grateful to my family. I would like to thank my mother Judy for her continued support throughout the years. Her love and encouragement continually provides me with the courage to reach for the stars. I also thank my father Robert for the many lessons along the way. I thank my brother Brett, for the constant reminder that life is full of beauty and endless possibilities. I am also deeply indebted to my grandparents, Herman and Jean. Their unconditional love and support has been more valuable than I could ever measure. I also thank my grandmother, my Lala, for reading every philosophy paper I have sent to her, for her inquisitiveness, and for our many discussions about Chalmers, consciousness, and life. I would also like to thank Kelly Haussler, whom I have always been able to rely on for both spiritual and emotional support.

I must also acknowledge Ian MacDonald, Aaron Warr, and Camille Campbell for their endless support and kindness. Finally, I am grateful to Jeff Sabine for his friendship and laughter during a difficult first year.

DEDICATION

For my Lala and Papa

TABLE OF CONTENTS

Introduction

Consciousness is a fascinating yet elusive phenomenon. While there is nothing we know more directly than our own consciousness, there also seems to be nothing about us that is harder to explain. Why does it exist? What function does it have? How could it possibly arise from neural or other processes in the brain and body? In the absence of a much more complete science of mind than we currently have, these questions remain unanswered, leaving consciousness as perplexing as ever.

Part of what makes consciousness so difficult to study is that it appears not to allow for a single agreed upon definition; rather, it is an ambiguous term that is used to refer to many different phenomena in a variety of ways (e.g., perception, thought, feeling, awareness, self-consciousness, etc.). Though it is necessary for a science of consciousness to account for each of these phenomena, some are easier to explain than others. For example, explaining how the visual system processes stimuli is relatively straightforward: as you continue to read this sentence, photons cause a conformational change in the retinol of your photoreceptors, which subsequently sends electrochemical signals up your optic nerve and into your visual cortex and other areas of your brain enabling you to react accordingly. Consciousness, however, is not comprised solely of information processing; it is also constituted by subjective experience. When you look at this page for example, you *experience* certain visual sensations: you have vivid impressions of the color, shape and size of the text. At the same time, you may begin to form thoughts or mental images, feel a given emotion, or have the urge to scratch an itch. These experiences are a part of your inner life, and trying to account for or explain them is one of the central problems involved with explaining consciousness.

To help clarify the questions and puzzles raised by consciousness, David Chalmers, a contemporary philosopher in the area of consciousness and cognitive science, distinguishes between the “easy problems” and the “hard problem” of consciousness. The easy problems of consciousness include explaining those phenomena that can be accounted for using the data collected with the standard methodologies of cognitive science (e.g., functional magnetic resonance imaging, positron emission tomography, computed tomography, etc.). The specification of appropriate computational and/or neural mechanisms is sufficient to explain

such phenomena, and thus, they do not pose a threat to a complete science of consciousness in the way that the "hard problem" does.

The easy problems include providing explanations of the following: How does the brain discriminate, categorize, and react to environmental stimuli? How does the brain integrate information from numerous sources and use this information to control behavior? How are human subjects able to access and report their mental states? These phenomena can be explained by accounting for the physical processes or mechanisms that correlate with them in the cognitive system. My previous example of the visual processing of stimuli is also an example of something currently easily explained via normal empirical scientific methods. And so, although more certainly needs to be understood about many of these phenomena, we have every reason to expect that cognitive science and neuroscience will continue to provide the details. Nevertheless, even a complete detailed understanding of the objective functioning of the brain would not be sufficient to solve the hard problem of consciousness, according to Chalmers.

The hard problem of consciousness is that of accounting for or explaining conscious *experience*. When we perceive, think, and feel, a whir of information-processing takes place in the brain, but this processing does not all go on in the dark; rather, it is often accompanied by states of experience. To use the phrase made famous by Thomas Nagel, *there is something it is like* to be a particular conscious individual at a particular time, having that individual's perceptual experiences, mental imageries, emotional experiences, and bodily sensations, and so forth.[1] For example, according to Nagel, there is something it is like to be me when I experience a visual sensation, such as that of a bright pink, or when I have an experience in another sense modality. Think about hearing the sound of a note being played on a piano, smelling the scent of a rose, or experiencing the pain associated with a deep sense of guilt: there is something it feels like to be in these states. Explaining how and why these states are accompanied by this subjective quality of experience is the central mystery of consciousness, and is Chalmers' hard problem.

[1] See Nagel, 436.

Unlike the easy problems, which can be explained away by specifying the neural or computational mechanisms involved in certain cognitive functions, Chalmers (and many other philosophers) believe that the hard problem would persist even after the performance of all relevant brain and bodily functions were explained.[2] The question of why all of this processing is accompanied by *experience* would still remain. Why is it that when my brain processes light of a certain wavelength, I have an experience of deep blue? Why do we have any experience at all? Merely identifying the mechanisms responsible for perceptual discrimination, motor function, language production, or emotional response does not tell us anything, or at least not enough, about the fact that we experience them.

But ultimately a complete theory of consciousness must explain *how* consciousness arises. If it arises from a physical system such as the human brain, *how* does this happen? Is consciousness itself physical, or is it perhaps an emergent property of a physical system? Though it seems likely that consciousness does arise from physical processes in the brain, we are clearly as yet unable to explain how. And no matter how much objective data we gather about the composition of the brain and the physical and chemical events that take place in it, we still seem to be left with something that we cannot explain; namely, how these generate subjective conscious experience.

A solution to the hard problem would thus involve an account of the relation between physical processes and consciousness, explaining how the physical processes are associated with states of experience. According to Chalmers, however, there exists an *explanatory gap* between the physical processes of the brain, and conscious experience itself. In order to bridge this gap, conscious experience cannot wholly be explained in terms of physical processes. Chalmers states that "[an] account of the functions stays on one side of the gap, so the materials for the bridge must be found elsewhere."[3] According to Chalmers' view of the mind, conscious experiences have irreducible, non-physical qualities, and therefore, any solution to the hard problem must be *non-reductive* (i.e., consciousness is non-physical and cannot be reduced to a physical system). Chalmers rejects the claim that the phenomenal aspect of

[2] See for example Nagel (1974), Jackson (1986), Broad (1925), and Feigl (1958).
[3] Chalmers, David. "Facing Up to the Problem of Consciousness" (1995), 206.

consciousness, that is, the subjective quality of experience, logically supervenes on the physical. On his view, then, our world is not exhausted by physical facts; rather, the presence of consciousness is an extra-physical fact about our world, and, furthermore, is one not guaranteed by the physical facts alone. And so, Chalmers argues that it is possible, in the sense that it is conceivable, that there could be worlds physically identical to ours but in which people, although otherwise entirely identical to people in the actual world, have no consciousness in the phenomenal sense. This claim is the core of Chalmers' conceivability argument against the view that consciousness can in principle be explained reductively. One of the problems associated with Chalmers' view however, is that it appears to make the emergence of consciousness, and of mental events in general, seem mysterious and inexplicable; something that a science of consciousness must necessarily avoid.

The overall objective of this thesis is to present, explore, and evaluate David Chalmers' argument that the existence of conscious experience constitutes a permanent barrier to the reductive aspirations of a purely materialistic neuroscience. Contra Chalmers, I will defend the possibility of a *reductive* explanation of consciousness, and argue that continued research in neuroscience and neurophysiology can result in a successful materialistic or reductive solution to the hard problem. My argument will be two-fold. First, I will challenge Chalmers' claim that consciousness does not logically supervene on the physical. And second, I will argue that his conceivability argument is implausible.

In presenting Chalmers' arguments, I refer primarily to his 1996 book *The Conscious Mind.* Though Chalmers has also provided detailed accounts of these arguments in other, in some cases more recent, works *The Conscious Mind* is Chalmers' most elaborate and influential work on this topic, and because his position has not changed since its publication, I see no problem using it as a primary source. [4]

[4] See for example "The Two-Dimensional Argument Against Materialism," "Phenomenal Concepts and the Explanatory Gap," Strong and Weak Emergence," "How Can we Construct a Science of Consciousness?" "Phenomenal Concepts and the Knowledge Argument," "Consciousness and Its Place in Nature," "Does Conceivability Entail Possibility?" and "Materialism and the Metaphysics of Modality."

CHAPTER ONE

THE IRREDUCIBILITY OF CONSCIOUSNESS

1.1 The Psychological and the Phenomenal Concepts of Mind

As cognitive science and psychology have demonstrated, conscious experience is not all there is to the mind. This shows us that two fundamental conceptions of the mind are often conflated; namely, the psychological and the phenomenal.[5] Cognitive scientists and psychologists are primarily interested in determining how mental states are responsible for the causation of behavior, and as such, are concerned with the *psychological conception of mind*. On this view of the mind, mental states are functional states that play the right sort of causal role in the production of behavior, or at least play an appropriate role in the explanation of behavior. It is irrelevant, on the psychological conception, whether a mental state has any conscious quality; all that matters is the role it plays in the cognitive system. In contrast, the *phenomenal conception of mind* is fundamentally concerned with the conscious experience aspect of the mental. On this view of the mind, a central thing that needs to be explained about mental states is how and why they feel, or are subjectively apprehended, as they are. This conception of the mind has been the primary interest of philosophers.

As we can see, on the psychological conception, the mind is characterized by what it *does* (i.e., its functional role in behavior), whereas on the phenomenal conception, the mind is characterized by how it *feels.* To make this distinction clear, imagine, for example, biting into a chocolate bar. A phenomenal investigation of the mind would attempt to explain why you experience the "chocolateness" or "sweetness" that you do and perhaps why you perceive

[5] Here, I follow Chalmers in distinguishing between the psychological and phenomenal conceptions of the mind. This contrast, however, is rather idiosyncratic. There is, after all, a discipline within psychology called phenomenological psychology. There are other ways of characterizing the mind and its aspects – in functional terms, or in terms of modularity, for example.

them differently than somebody else. A psychological investigation, on the other hand, would concern itself with the neural circuits that are activated during the bite, establishing how these circuits might be represented computationally, and determining how neural processing taking place explains when you might stop eating. Neither the psychological nor the phenomenal conception provides *the* correct analysis of the mind; rather, they focus on different aspects of mental phenomena (e.g., functional and experiential) and are both necessary for a complete understanding of the mind.

A given mental concept, one that refers to a mental state, can then be understood as a psychological concept, a phenomenal concept, or as a hybrid of the two. Take, for example, the mental concepts of itchiness and learning. Itchiness has certain conscious, or qualitative, feels associated with it, without which we would not identify it as itchiness, and it is therefore perhaps best understood as a phenomenal concept. In comparison, the concept of learning might best be taken as a psychological concept since for something to learn is for it to appropriately adapt its behavioral capacities in response to certain kinds of environmental stimulation.[6] However, most mental concepts straddle the fence, having both phenomenal and psychological components. Take, for example, the sensation of pain. In our ordinary language, "pain" is most commonly used to refer to a certain class of unpleasant qualitative feels. When used this way, "pain" would best be understood as a phenomenal concept. However, there is

[6]Here it can be argued that learning is not exclusively a psychological (or functional) concept, since *there is something it feels like* to learn. However, what is central in an explanation of learning is determining how a cognitive system adapts to various new circumstances and stimuli. While there may be a phenomenal quality present during parts of learning, it is not required for a particular cognitive process to count as an instance of learning. To help determine whether a mental concept is primarily psychological or phenomenal, Chalmers formulated a simple test:

> "A good test for whether a mental [concept] *M* is primarily psychological is to ask oneself: Could something be an instance of *M* without any particular associated phenomenal quality? If so, then *M* is likely psychological. If not, then *M* is phenomenal, or at least a combined notion…" (1996, 18).

While it is possible for learning to have an associated quality of experience, it can also be implicit or unconscious (i.e., learning that occurs without the intention to learn and largely or completely without awareness of the nature of what has been learned), and as such, not have an associated quality of experience. In general, wherever there is phenomenal consciousness there is awareness; this will be discussed in detail below. We can see then, that learning occurs regardless of whether there is an associated phenomenal quality present. What is central to the concept of learning, therefore, is its playing of an appropriate cognitive role, and as such, it is best understood as a psychological concept.

also a psychological feature associated with the term: roughly, pain is a state produced when a person suffers some form of damage, and which tends to lead to aversion reactions. Both the phenomenal and psychological aspects of pain are necessary for a complete understanding of the concept. Likewise, to have a complete understanding of the concept of perception generally, both phenomenal and psychological aspects must be recognized. Taken phenomenally, perception refers to the conscious quality of perceptual experience. Taken purely psychologically, however, perception denotes a process whereby a cognitive system is sensitive to environmental stimulation in such a way that the resulting states play a certain role in directing cognitive processes.

"Consciousness" itself is also a hybrid term, having both phenomenal and psychological aspects. In the phenomenal sense, consciousness refers to the qualitative feels (i.e., intrinsic qualities) of experience. In the psychological sense, however, consciousness can be used to refer to a number of psychological properties, including awakeness, reportability, and voluntary control. Among other things, we often say that a person is conscious when they are awake (i.e., not asleep), when they have the ability to answer questions about their mental states, and when they have control over their behavioral acts (i.e., their actions are deliberate). These psychological aspects can be explained in terms of their functional roles. Hence, "awakeness" refers to the ability to process information about the world and react to it appropriately, "reportability" refers to processes whereby information about internal states is retrieved and made available for verbal report, and "voluntary control" refers to the ability to deliberately perform an action, where the action is caused by an earlier state of the cognitive system. Although these concepts might best be understood as psychological concepts, they nevertheless all typically have phenomenal qualities associated with them. Therefore, in order to fully understand consciousness, we must recognize that it not only has a phenomenal aspect, but that there are a number of psychological concepts associated with the term, and each of those concepts also refer to entities or processes with phenomenal qualities. To avoid confusion, let us refer to the psychological aspects of consciousness as *psychological consciousness,* and the phenomenal aspect as *phenomenal consciousness.*

According to Chalmers, there is at least one psychological property associated with phenomenal consciousness; namely, awareness.[7] Awareness denotes a process wherein we come to have access to some information, and can use that information to direct our behavior.[8] In general, where there is phenomenal consciousness, there will also be awareness. For example, when I have a phenomenal experience of pain, it is accompanied by my awareness of something unpleasant, and will lead (if possible) to aversion behavior. For Chalmers, we know that conscious experience is accompanied by awareness because our experiences are reportable (at least potentially). That is, even in cases where we are not paying attention to an experience, or some aspect of it, we have the ability to shift our attention to it, thereby making it reportable. Although phenomenal consciousness and awareness co-occur, they are conceptually distinct.

As we can see, it might often be difficult to differentiate between phenomenal and psychological properties because they tend to co-occur. According to Chalmers, this co-occurrence is nothing more than a mere correlation (i.e., it does not imply an ontological equivalence or dependence).[9] When a person experiences pain, the phenomenal qualities associated with the experience are accompanied by psychological processes; however, we cannot reduce the phenomenal pain to these psychological processes because the phenomenal and psychological are distinct concepts referring to distinct phenomena requiring distinct explanations. At this stage, I do not want to beg the question about whether the phenomenal and psychological will turn out to be the same thing, or that the phenomenal can indeed be reduced to the psychological. For now, what is important is recognizing the conceptual distinction between the two notions: what it means for a mental state to be phenomenal is for it to feel a certain way, and what it means for a mental state to be psychological is for it to play an appropriate causal role.

---

[7] Chalmers (1996), 28.

[8] It should be noted that Chalmers focuses solely on the *conscious* processing of awareness (see Chalmers (1996), 28). Presumably, however, there are *unconscious* processes that are also involved in awareness. This serves as another example of Chalmers' idiosyncratic usage of terms (see also page 7, footnote 5).

[9] Here, I mean to include causal dependence as a form of ontological dependence.

It should be noted that awareness does not entail the ability to report. Although animals and infants lack the capacity for language, we would not want to deny that they have awareness (at least to some degree). After all, if I were to accidently step on a cat or infant, both would likely respond by crying out in pain. This behavioral response would indicate that they are aware of something unpleasant having just occurred, even though they presumably lack the concepts and linguistic ability to make what we would count as a report. From this point, whenever I use the term "consciousness" without qualification I am using it to refer exclusively to phenomenal consciousness. When I wish to speak of the psychological aspects of consciousness, I will refer to "psychological consciousness."

1.2 Historical Foundations

The conflation of phenomenal and psychological aspects of the mind has a long history. In 1641, René Descartes was one of the first modern philosophers to make this mistake. Descartes held that the mind is transparent to itself, and that every mental event is a *cogitatio,* or a content of experience. As such, Descartes identified the mental with the phenomenal. In his reply to the Fourth Set of Objections, he writes:

> As to the fact that there can be nothing in the mind, in so far as it is a thinking thing, of which it is not aware, this seems to me to be self-evident. For there is nothing that we can understand to be in the mind, regarded in this way, that is not a thought or dependent on a thought. If it were not a thought nor dependent on a thought it would not belong to the mind *qua* thinking thing; and we cannot have any thought of which we are not aware at the very moment it is in us.[10]

We see in this quote that Descartes failed to recognize the distinction between the two aspects of the mind, and instead, classified all mental states (including psychological states) as having phenomenal properties.[11] Although Descartes did not separate the phenomenal from the psychological, he did set the stage for a more broad inquiry into understanding how the mind interacts with the body.

[10] Descartes (1984), 156.

[11] It is interesting to note that Descartes often excluded sensations from the category of the mental, and instead classified them as being a phenomenon associated with the body (i.e., the physical). For Descartes, therefore, not only were psychological states assimilated to the phenomenal, but not every phenomenal state (at least as I am understating the notion) would be included in the realm of the mental.

In his *Meditations on First Philosophy,* Descartes reflects on the mind and the body and concludes that they are fundamentally distinct from one another. That is, he argued that minds and bodies are composed of different substances and have different attributes or modes (i.e., mental and physical). Descartes' substance dualism holds that the mind is made of some immaterial substance, while the brain (along with the rest of the human body) is composed of a different substance, matter. The majority of philosophers, however, agree that substance dualism has the serious difficulty that the mind cannot affect the brain (or body), because of the ostensible impossibility of an immaterial thing, something that takes up no space, has no mass, or other properties that concrete material objects have, interacting causally with a physical and extended thing, the brain.[12] This problem is even more difficult if we not only accept that mental states and brain states have distinct natures, but also maintain that mental states are not reducible to physical processes.

It was not until the late nineteenth century that Wilhelm Wundt, a German psychologist, put forward his *principle of actuality* in opposition to Descartes' substance dualism. According to this principle, all mental content is an *event* or *activity* of the brain. Wundt agreed with Descartes that the nature of the mental is fundamentally distinct from that of the physical, however, unlike Descartes, who believed that the mental and the physical interacted through the pineal gland, Wundt maintained that it is inconceivable that these two realms could causally influence each other. And so, Wundt claimed that there must be two causal chains that run in parallel, the chain of mental causality and the chain of physical causality. On this view, mental events can only be caused by mental events, and physical events can only be caused by physical events, it is just the case that certain mental events regularly co-occur with certain physical events.[13] Keep in mind, however, that "regular co-occurrence" does not imply causation; at most, it implies that there is a *correlation* between mental events and

[12] Though Descartes believed that the nature of the mind and body were fundamentally distinct, he was an interactionist (i.e., the view that the mental and physical are fundamentally distinct but interact in both directions – physical states affect mental states, and mental states affect physical states). According to Descartes, the point of interaction between the mind and body was the pineal gland (see Descartes (2002), 22). However, the idea that the pineal gland mediates this interaction has long been rejected on both physiological and philosophical grounds.
[13] Kusch, 134.

physical events. By viewing the mental as an activity rather than as a substance, Wundt seemingly avoids Descartes mind-body problem.

Wundt was one of the first people to differentiate between the two aspects of the mind. Interested in applying the methodological and experimental methods of science to the study of consciousness, Wundt established the first laboratory of experimental psychology. Participants in Wundt's experiments were exposed to a range of sensory stimuli (e.g., auditory, visual, tactile, olfactory, etc.) and then asked to report their inner experiences via introspection (i.e., the process of looking within oneself to observe one's own conscious experiences).[14] The school of thought that arose from the work of Wundt and his colleges is called *structuralism.* The structuralists' primary objective was to study consciousness by breaking it down into what they took to be its basic components (i.e., sensation, perception, and affection).

Structuralism eventually gave way to an opposing school of thought, *functionalism,* led by psychologist William James, and heavily influenced by Darwin's evolutionary theory. Functionalists were interested in determining the purpose, or function, of the mind as opposed to merely analyzing its structure. For James, consciousness is an ever-changing stream or flow of images and sensations. And so, rather than studying the mind from the standpoint of its composition (i.e., its mental elements of structure), functionalists see it as an accumulation of processes that have helped us adapt and survive in our environment. In other words, functionalists are concerned with how the mind processes information and directs behavior.

In the 1920's, structuralism and functionalism were challenged by *behaviorism,* a school of thought that emphasized the study of observable behavior.[15] Behaviorist John Watson believed that in their study of the mind, both structuralism and functionalism diverged too much from objective science. Watson was strongly opposed to the study of the mind and of conscious experience, and argued that the proper subject matter of psychology was observable behavior.[16] Watson and the behaviorists, therefore, believed introspection to be inappropriate as a scientific method because it relied on a subject's unobservable inner states for data. Psychology, on this view, was the science of behavior, not the science of the mind. Behaviorists

[14] Passer, 14.
[15] Passer, 21.
[16] Passer, 21.

held the source of behavior to be external (i.e., in the environment) as opposed to internal (i.e., in the mind), and claimed that it can be explained without making reference to mental events or to internal psychological processes. Although most behaviorists did not deny the existence of mental states, they considered them irrelevant to a scientific explanation of behavior. Hence, behaviorism established the idea that a psychological explanation of the mind should proceed with phenomenal consciousness ignored.

In analytic philosophy, on the other hand, the shift in emphasis from a phenomenal to psychological understanding of the mind was definitively established in 1949, when Gilbert Ryle argued that mental states can be analyzed in terms of dispositions to behave in certain ways.[17] This view, *logical behaviorism,* holds that the mind is an aspect of behavior and should not be seen as distinct, or separate, from the body. To be in a given mental state therefore, is to be in a certain behavioral state, or at least to have a disposition to behave in a certain way. In this sense, the mind has an observable public aspect (i.e., behavior) as opposed to only a strictly inner aspect. However, it seems counterintuitive to claim that the mind is an *aspect* of behavior rather than an inner *cause* of behavior. That is, it is not clear that mental states are reducible to behavioral states (or behavioral dispositions). After all, it is possible that an actor might have a given behavioral disposition without the associated mental state. If so, mental states cannot be behavioral dispositions.

The more recent philosophical doctrine of *functionalism* emerged as a way to resolve some of the problems presented by logical behaviorism.[18] In general, functionalism holds that mental states correspond to functional states (i.e., states that play a certain role in the cognitive system). Advocates of this view, including David Lewis (1966) and David Armstrong (1968), maintain that mental states are defined in terms of their causal role; that is, in terms of the kinds of stimulation that tend to produce them, the kind of behavior they tend to produce, and the way they interact with other mental states.[19] Contra logical behaviorism therefore,

[17] See Ryle (1949).
[18] The philosophical doctrine of functionalism should not be confused with the functionalism of William James.
[19] There are other forms of functionalism, including Hilary Putnam's machine functionalism (1960). However, I do not consider this alternative here because it is an empirical hypothesis that is not concerned with the analysis of mental concepts.

functionalism maintains that mental states are internal states that stand in the right kind of causal relation to behavior. In this sense, functionalism overcomes our previous concerns with logical behaviorism. However, not *all* mental concepts seem amenable to being analyzed in this way (in particular, phenomenal concepts). For example, if I were to play a song for you on the piano, I would likely not care whether you were receiving environmental stimulation and processing it in a certain way, but rather, whether you were *experiencing* an auditory sensation. Determining the function, or causal role of a given mental state, does not answer why there is a phenomenal quality associated with that state. Functionalism, therefore, also seems to be guilty of ignoring the phenomenal aspect of experience.[20]

As these brief sketches from the history of psychology and philosophy demonstrate, separating the phenomenal from the psychological aspects of the mind has not been easy. While this discussion is in no way a complete examination of the historical difficulties surrounding the two conceptions, or their history and interplay, it makes clear that the psychological and the phenomenal are both real and distinct aspects of the mind.

What is also important to recognize is that psychological concepts deal with aspects of the mind that are third-person accessible, and therefore arguably objective, while phenomenal concepts refer to phenomena that are primarily, and perhaps solely, first-person accessible, and therefore arguably subjective. Third-person data is empirically based, and so, is dependent on evidence that is observable by the senses. In the case of consciousness, therefore, third-person data can include information about behavior and brain processes. On the other hand, only first-person data can report the phenomenal quality of conscious experience which is not directly observable from the third-person perspective. First-person data can include information about emotional experiences, mental imagery, and occurrent thought (e.g., the experience of reflecting and deciding). Chalmers argues that first-person data is irreducible to third-person data.[21] That is, first-person data about phenomenal consciousness cannot be explained in terms of third-person data about brain processes. On Chalmers' view, therefore, a complete theory of consciousness must reconcile the objective (i.e., psychological) aspects of

[20] It should be noted, however, that there are a number of functionalists who argue that the phenomenal aspect of experience can be captured in functional terms (see Shoemaker (1975) and White (1986)).
[21] Chalmers (1996), 93.

consciousness with the subjective (i.e., phenomenal), but cannot do so by simply reducing the subjective to the objective.

1.3 First Person vs. Third Person Data

The notions of the psychological and phenomenal aspects of mental states, and the different ways in which we are able to access them, are closely connected to Chalmers' distinction between the hard and easy problems of consciousness. For Chalmers, the easy problems of consciousness are solvable with data generated from purely the third-person perspective, while, on the other hand, first-person data is necessary to solve the hard problem. This is because the hard problem requires explanation of the phenomenal aspect of the conscious mind, not the psychological. Even if all the psychological structures and functions of the mind were explained, we would still have subjective experience, which remains as an extra ingredient. In other words, third-person data can explain the objective functioning of a system and can ultimately be scrutinized by external assessment. However, this kind of scrutiny is not possible when dealing with subjective inner states.

Explaining observable phenomena with third person data is something we know how to do, and have the language and concepts to accomplish. For example, to explain perceptual discrimination, we need to explain how some cognitive processes distinguish between various stimuli and produce appropriate responses. In order to do so, we need to identify the neural mechanism(s) responsible for distinguishing the relevant stimuli. Neuroscience is responsible for determining how the brain integrates information from various areas of the cortex in order to initiate a behavioral response. In many cases, neuroscientists are capable of physically manipulating specific regions of the brain in order to detect their functions, as well as to verify which neural correlates are responsible for specific actions.

On the other hand, experiences such as pain, desire, hunger, color perception, and mental imagery, if we consider them purely phenomenologically or in terms of their experiential aspects, are internal occurrences with a first-person nature that, according to Chalmers, cannot be directly or fully assessed by external observers. Nevertheless, it seems clear that an anesthesiologist can obtain a general idea of an individual's threshold for pain, based on third-person, empirical evidence. While it is impossible for an anesthesiologist to

physically peer into a person's brain and see the pain that they are experiencing, there are kinds of evidence available that allow anesthesiologists to identify how much stimulus of a certain sort can be imposed before the intensity becomes intolerable to the patient. If this evidence is independent of a patient's verbal reports, then there would be reason to believe that mental states are correlated, if not to brain states, then at least to some physical state.[22] This is a problem for Chalmers, since he claims that conscious experience cannot be analyzed through objective methods.

As the anesthesiologist example shows, there seems to be (at least in some cases) a systematic association between first-person data and third-person data regarding consciousness. For Chalmers, being able to integrate these two frameworks is essential to the understanding of consciousness as a whole. But, while Chalmers acknowledges that there is a correlation between the two forms of data, he also claims that this correlation is not equivalent to, or sufficient for, explanation of the phenomenal. Chalmers argues that a science of consciousness must proceed by analyzing the relationship between first-person and third-person data without resorting to reductionism. In other words, phenomenal 'feels' cannot be reduced to or explained by mere physical events or processes.

In order to be considered a legitimate and scientific theory, a complete theory of consciousness must be constructed from data that is acceptable to scientific methods (e.g., publically verifiable, repeatable, and measurable). Under Chalmers' view, however, even if mental states and brain states are correlated, the phenomenal aspects of mental events cannot be publically verified. Currently, the only way to access or identify a person's inner states is through their introspection and reporting. The problem, however, is that introspection is not a valid scientific method because the data it produces is not reliable (e.g., it cannot be verified by objective methods). There is no way to verify whether a person's verbal report about their experiences accurately reflects the contents of their experience. Moreover, it seems introspective data cannot be replicated, since each subjective experience is arguably a singular, and thus non-repeatable, event. This makes introspection problematic as a method for producing generalizations across individuals, or even within individuals across time. And so,

[22] I will discuss the problems associated with introspective reports later in this section.

introspection is not an acceptable approach to developing a scientific theory of the mind. The predicament Chalmers has identified now becomes clear.

Although neuroscientists are capable of viewing the inner workings of a person's brain, if phenomenal states are not brain states, then it is impossible, via merely third person data, to identify someone's subjective experiences. However, as my discussion concerning the anesthesiologist shows, there seem to be at least some cases in which a person's subjective experiences are available for third-person analysis, even if only indirectly. Nevertheless, in order to obtain a complete theory of the mind, it is necessary that *all* subjective experiences be explained (or in principle explicable) using objective methods. According to Chalmers, however, neuroscience will never be able to provide a full account of conscious experience because conscious experience is a "fundamental feature [of the world], irreducible to anything more basic," (like brain states or processes).[23] And so, something over and above brain states must be provided or invoked in order to account for phenomenal consciousness.

Chalmers' view that phenomenal properties are ontologically independent and distinct from physical properties constitutes a kind of *property dualism*: conscious experience involves properties of an individual that are of a different kind than their physical properties, and neither is reducible to the other. For Chalmers, therefore, "consciousness is a non-physical feature of the world."[24] While I will discuss in more detail Chalmers' argument for this claim in the next section, the basic idea is that the physical structures of the world (e.g., the exact distribution of particles, fields, and forces in space-time) are "logically consistent with the absence of consciousness, so the presence of consciousness is a further fact about the world."[25] In other words, although phenomenal consciousness is a fact about the world, it cannot be reduced to a physical process; instead, consciousness exists over and above any physical or neurological process.

According to Chalmers, "it is possible that consciousness *arises* from a physical basis, even though it is not *entailed* by that basis."[26] Therefore, consciousness does not *logically*

[23] Chalmers. "The Puzzle of Conscious Experience." (1995), 96.
[24] Chalmers (1997).
[25] Chalmers (1997).
[26] Chalmers (1996), 126.

supervene on the physical. The physical features and laws of the world are, on Chalmers' view, not sufficient to explain consciousness. Since consciousness is a fundamental feature of the world (and fundamental features cannot be explained in terms of more basic features), Chalmers argues that consciousness must simply be taken as ontologically primitive. Clearly then this position entails that phenomenal consciousness is rather mysterious; it cannot be explained by any physical properties or processes but is simply somehow part of the basic furniture of the universe. Imagining how the sciences could get a handle on explaining it thus appears pretty difficult if Chalmers is correct about the status of phenomenal consciousness. Putting that aside for now, let us explore Chalmers' argument against a reductive explanation of consciousness in more detail. At the center of his argument is the notion of supervenience. For Chalmers, in order for a reductive explanation of consciousness to be possible, it has to not merely arise from physical systems and their properties, but be "globally logically supervenient on physical properties."[27] But what does this mean?

1.4 Supervenience

Clearly, Chalmers' view contrasts with both materialism and physicalism. And although Chalmers does not differentiate between materialism and physicalism, the distinction between them is an important one. While materialism and physicalism are both materialist theories (i.e., they both hold that everything is in some sense ultimately constituted by matter), physicalism is a somewhat broader metaphysical position which evolved with the natural sciences to include phenomena not necessarily precisely material (e.g., gravity, wave/particle relationships, energy, etc.). For example, although we cannot see gravity, we know, based on physics, that gravity is a real thing in the universe and the result of certain physical properties holding. Thus, some phenomena, like gravity, which cannot be straightforwardly reduced to physical phenomena are nevertheless physical, at least in a broad sense, because they are supervenient on, are dependent upon, or emerge from, physical phenomena. Since he explicitly denies that phenomenal qualities are *themselves* physical, I take it that Chalmers' argument is really against physicalism about the mental (though his rejection of physicalism implies the rejection of materialism as a whole).

[27] Chalmers (1996), 48.

The notion of *supervenience,* the idea that one set of facts can fully determine another set of facts, is central to physicalism, and thus is also central to Chalmers' argument against a reductive explanation of consciousness. This makes it important to understand Chalmers' view of the nature of supervenience. Chalmers' definition of supervenience is:

> "B-properties supervene on A-properties if no two possible situations are identical with respect to their A-properties while differing in their B-properties."[28]

In other words, B-properties supervene on A-properties if, for any two possible situations, if both situations are identical concerning A-properties, they are also identical concerning B-properties.[29] Hence, taking biological properties as an example, we would say that biological properties supervene on physical properties insofar as any two possible situations that are physically identical are also biologically identical. However, this general definition serves only as a template; more precise notions of supervenience can be obtained depending on how we fill it in.

For example, within this general definition, "situations" could be used to refer to individuals or entire worlds. If used to refer to individuals, we obtain the notion of *local* supervenience, and if used to refer to entire worlds, we obtain the notion of *global* supervenience. Moreover, depending on how we construe "possibility," we generate notions of logical, natural, or other kinds of supervenience. Chalmers' arguments involve only *logical* and *natural* supervenience, but the distinction between global and local also plays some role. I'll begin by explaining the difference between local and global supervenience.

1.5 Local and Global Supervenience

Local supervenience is concerned with properties as they relate to individuals. Chalmers defines it as follows: "B-properties supervene *locally* on A-properties if the A-properties of an *individua*l determine the B-properties of that individual"[30] For example, shape supervenes locally on physical properties, since any two objects with the same physical properties will

[28] Chalmers (1996), 33.

[29] Here, I use "identical" in the sense of qualitative identity. An object *x* and an object *y* are *qualitatively* identical if *x* and *y* do not differ in respect to their qualities - they have exactly similar qualities as one another (e.g., identical twins). In contrast, *x* and *y* are *numerically* identical if they are one and the same thing (i.e., they are not two distinct things).

[30] Chalmers (1996), 33.

necessarily have the same shape. However, local supervenience of a property on the physical fails if that property is somehow context-dependent (i.e., if an object's possessing that property depends not only on the object's physical constitution, but also on its history or environment). For example, the aesthetic value of the Mona Lisa might supervene on the physical composition of the painting (e.g., the specific particles that make up the painting). However, because we could have an exact physical replica of the Mona Lisa, we would not say that the Mona Lisa's aesthetic value locally supervenes on its physical composition, because (arguably) the aesthetic value of the original is also dependent on its history (e.g., the fact that it was painted by Leonardo da Vinci).

In contrast, *global supervenience* is concerned with properties as they relate to entire worlds. And so, B-properties supervene *globally* on A-properties if, for any two *possible worlds*, w and v, if w and v are identical concerning A-properties, they are also identical concerning B-properties. For example, if biological properties supervene globally on the physical properties of our world, any world physically identical to ours will also be biologically identical. However, if biological properties don't supervene globally, but only locally, it is possible to have two physically identical people that nevertheless differ in certain biological characteristics. For example, due to environmental differences, one of them might be biologically more fit (i.e., better suited for reproductive success).

As we can see, global supervenience allows for B-properties to be determined not just by local properties of an individual, but by some wider spatiotemporal distribution of properties. However, for Chalmers, when it comes to conscious experience, the distinction between local and global supervenience does not matter too much, since he believes that "it is likely that insofar as consciousness supervenes on the physical at all, it supervenes locally."[31] Nevertheless, it is crucial to recognize that on Chalmers' view, a reductive explanation of consciousness is possible only if consciousness globally supervenes on the physical not naturally, but logically. And so, for our purpose, a more important distinction is between logical (or conceptual) supervenience, and natural (or empirical) supervenience.

[31] Chalmers (1996), 34.

1.6 Logical and Natural Supervenience

According to Chalmers, "B-properties supervene *logically* on A-properties if no two *logically possible* situations are identical with respect to their A-properties but distinct with respect to their B-properties."[32] Here, logical possibility should be understood in a broad sense. That is, in determining what counts as a logically possible situation, the constraints are largely conceptual. And so, a logically possible situation is one that contains no internal inconsistencies or contradictions. For example, a male vixen is logically impossible: the concept of being a male is in contradiction with the concept of a vixen (since by definition a vixen is a female fox). On the other hand, the notion of a flying telephone is conceptually coherent (i.e., taken together, the concept of flying and the concept of a telephone are not contradictory) and therefore flying telephones are logically possible. Logical possibility also constrains God (hypothetically). For example, God could not have created a world with married bachelors (since, by definition, "bachelor" refers to an unmarried male), but could have created one where the cow did, in fact, jump over the moon. As I mentioned above, the notion of logical possibility is central to Chalmers' argument. This is because, put simply, Chalmers' argument against a reductive explanation of consciousness can be formulated in the following way:

(1) Consciousness can be reductively explained only if it is *logically* supervenient on the physical.

(2) Consciousness is *not* logically supervenient on the physical.

Therefore, (3) Consciousness is *not* reductively explainable.

One possible problem for Chalmers' view arises, however, in his explanation of logical supervenience, since Chalmers uses the term in two distinct ways: as an ontological relation of determination and as an epistemological relation whereby knowledge of the A-facts (i.e., facts about A-properties) entails knowledge of the B-facts (i.e., facts about B-properties).[33] Chalmers does not differentiate between these two interpretations, and instead treats them as equivalent. In chapter two, I will argue that these interpretations are not equivalent, and that Chalmers is thus guilty of equivocation. The plausibility of premise (1) rests on an ontological interpretation of logical supervenience, while that of premise (2) rests on an epistemological

[32] Chalmers (1996), 35.

[33] This criticism is, in part, credited to Rowlands (2001).

interpretation. For now, however, let us assume that Chalmers does not make this mistake so that we may see how his argument (as he conceives it) is supposed to unfold.

Chalmers explains logical supervenience in terms of what God would have to do in order to create a world containing certain facts. For Chalmers, B-facts logically supervene on A-facts if it is the case that once God creates the A-facts, the B-facts are automatically established. For example, at a global level, biological properties supervene *logically* on the physical. And so, on the divine explanation of logical supervenience, God could not have created a world that was physically identical to ours but biologically different, for there is simply no logical space for the biological facts to vary independently of the physical facts. Here, the B-facts are, in an important sense, constituted by the A-facts.[34]

Moreover, Chalmers also uses the notion of a hypothetical super-being to explain that when logical supervenience holds, the only information necessary for coming to know all the biological facts is the microphysical facts of the universe. The super-being, if she knew the exact location of every particle in the universe at every point in time, and therefore all the microphysical facts of the universe, would be able to determine the biological facts. Hence, the microphysical facts are sufficient for such a being to determine the microscopic structure and dynamics of the world, which, in turn, is sufficient information for her to straightforwardly deduce the macroscopic biological structures and dynamics.[35] Thus, we can say that B-properties supervene logically on A-properties when the A-facts entail the B-facts. But recall that for Chalmers' anti-physicalism regarding consciousness, the distinction between logical and natural supervenience is crucial.

According to Chalmers, "B-properties supervene *naturally* on A-properties if any two *naturally possible* situations with the same A-properties have the same B-properties."[36] Unlike logical possibility, natural possibility encompasses only situations that could occur in the actual world, in the sense that they do not violate the natural laws that hold here. Natural possibility thus imposes a much stronger limitation than mere logical possibility. For example, although a universe without gravity is logically possible (i.e., we can coherently conceive a universe that is

[34] In chapter two, I will refer to this as the ontological interpretation of logical supervenience.
[35] In chapter two, I will refer to this as the epistemological interpretation of logical supervenience.
[36] Chalmers (1996), 36.

not constrained by the laws of gravity) it is not naturally possible (i.e., given our natural laws, gravity must exist). Naturally possible situations, therefore, are ones that conform to the laws of nature of our world.

Natural supervenience holds when, among all naturally possible situations, the A-facts necessitate the B-facts. This occurs when specific A-properties in our world are continually accompanied by specific B-properties, and when this correlation is not just coincidental, but lawful. For something to count as a law, however, it must hold not just in actual and current situations but also in counterfactual and future situations. And so, the dependence of B-facts on A-facts must support counterfactual situations – that is, supervenience holds only if it is the case that if the A-facts did not occur then the B-facts would not have occurred – as well as any situations that could potentially occur in the future – that is, whenever the A-facts actually occur the B-facts will also occur.[37] Although this co-occurrence need not hold in every logically possible situation, it must hold in every naturally possible situation.

The way Chalmers explains the difference between logical supervenience and mere natural supervenience can be thought of in the following way: if B-properties supervene *logically* on A-properties, then once God created a world consisting of certain A-facts, the B-facts would thereby be established automatically. If B-properties merely supervene *naturally* on A-properties, however, then after creating the A-facts, God would have more work to do in order that the B-facts would also be established; that is, She would have to establish laws relating the A-facts and the B-facts. For Chalmers, it is clear that logical supervenience implies natural supervenience: if any two logically possible situations with the same A-properties have the same B-properties, then the same will be true of any two naturally possible situations, because naturally possible situations are a subset of logically possible situations. However, as we will see, natural supervenience does not imply logical supervenience.

According to Chalmers, consciousness is naturally supervenient on physical properties insofar as in the actual world, any two physically identical individuals in physically identical

[37] It is possible for there to be cases where a given B-fact can supervene on different sets of A-facts. That is, a given B-fact may supervene on A-fact$_1$, A-fact$_2$…A-fact$_n$. For example, a qualitative pain state might be identical in the case of two or more different physiological substrates. In these types of cases, we would say the B-facts hold only if the A-facts, or A-facts$_2$, or A-facts$_n$ hold; i.e. if the B-facts supervene on the disjunction of these A-facts.

states will have qualitatively identical experiences. He claims however that "it is not at all clear that consciousness is logically supervenient on physical properties."[38] Chalmers argues that it is *logically* possible (i.e., conceivable) for a being physically identical to a conscious human being to not have conscious experience at all (or to have phenomenally different conscious experiences). In the philosophic literature, these claimed beings without phenomenal consciousness are called "zombies." Chalmers' central argument against physicalism is that the conceivability of zombies demonstrates that consciousness supervenes only naturally and not logically on the physical. According to him, the connection between physical structures or processes and consciousness therefore is established by the laws of nature, and not by any logical, or conceptual, force or connection. Chalmers' assertion of the conceivability of zombies shows that consciousness cannot be reductively explained. This is Chalmers' conceivability argument. Before explaining this argument in a bit more detail, I will briefly address a problem that Chalmers acknowledges with the notion of logical supervenience as he has so far defined it.

1.7 A Problem with Logical Supervenience

As Chalmers notes, there is a problem with the notion of logical supervenience as he has defined it thus far. This problem arises from the logical possibility of a world that is physically identical to ours, but which has additional non-physical properties that are not present in our world. Using Chalmers' example, we can conceive a world that is physically identical to the actual world, but which has angels hovering in a non-physical realm also part of that world.[39] If these angels were capable of reproducing they would have their own evolutionary history (assuming they were not in an unchanging environment to which they were perfectly adapted). And so, the angel world would be physically identical to the actual world, but biologically distinct, because there would be biological facts in the angel world that are not the case in ours. Under Chalmers' current definition of logical supervenience, this implies that biological properties are not logically supervenient on physical properties. However, in the actual world, biological properties *are* logically determined by the physical facts. For Chalmers, it is

[38] Chalmers (1996), 37.

[39] Chalmers (1996), 39. Here, I follow Chalmers' assumption that there are no angels in the actual world.

objectionable that the "mere logical possibility of the angel world stand in the way of the determination of biological properties by physical properties in our own world."[40]

In order to retain a notion of logical supervenience that is compatible with the possibility of an angel world (or the possibility of any other world that is composed of extra, non-physical properties), Chalmers must modify his original definition. This modification is as follows:

> "B-properties are logically supervenient on A-properties if the B-properties in our world are logically determined by the A-properties in the following sense: in any possible world with the same A-facts, the same B-facts will hold."[41]

Under this revised definition, biological properties logically supervene on physical properties in the actual world, despite the existence of possible worlds where this is not the case. Hence, the existence of possible worlds with *extra* B-facts will not affect the supervenient relation between A-facts and B-facts in the actual world. The only thing that matters, according to Chalmers, is that the B-facts that hold in the actual world also hold in any physically identical world. Now that we have a more accurate understanding of how Chalmers defines the notion of logical supervenience, let us turn, once again, to his conceivability argument.

1.8 Chalmers' Conceivability Argument

According to Chalmers, my zombie twin is a being who is physically, functionally, and psychologically identical to me, but who is nevertheless phenomenally impaired, indeed vacuous. My zombie twin lacks conscious, that is, phenomenal experience. Although my zombie twin is psychologically conscious, in the sense that she is awake, able to report the contents of her internal states, and able to shift the focus of her attention, none of this functioning is accompanied by phenomenal experience. There is nothing it feels like to be my zombie twin (or any zombie in general).

Chalmers' conceivability argument is thus that the logical possibility of zombies demonstrates that there is no logical entailment from physical facts to facts concerning phenomenal consciousness. And so, the existence of consciousness is a further, non-physical,

[40] Chalmers (1996), 39.
[41] Chalmers (1996), 39.

fact about the world. Although Chalmers acknowledges that zombies are naturally impossible – inconsistent with the natural laws of the actual world – he nevertheless claims that they are logically possible (i.e., conceivable). The conceivability of philosophical zombies is Chalmers' support for the claim that consciousness is not identical to, realized by, or supervenient on physical properties. Since zombies are physically identical to conscious human beings but lack phenomenal conscious experience, their possibility points to an epistemic gap between our conception of the physical facts of the world, and our conception of consciousness. Chalmers' argument can thus also be formulated as follows, where P is a complete physical description of the world, including the fundamental physical laws, and Q is the existence of phenomenal consciousness (~Q, therefore, is the non-existence of phenomenal consciousness):[42]

(1) P & ~Q is conceivable.

(2) If P & ~Q is conceivable then P & ~Q is metaphysically possible.

(3) If P & ~Q is metaphysically possible then physicalism is false

Therefore, (4) Physicalism is false.

Furthermore, conceivability is not simply a matter of whether a given situation can be imagined, but rather, conceivability is connected to the nature of physical and phenomenal concepts. A statement is conceivable if it is consistent with the totality of conceptual truths, where conceptual truths are truths in virtue of meaning (i.e., a statement is true simply by virtue of the meanings of the terms involved). Moreover, it has been traditionally assumed that all conceptual truths are knowable *a priori* (i.e., knowable in the absence of any empirical investigation or experience). And so, to support the premise that zombies are conceivable, Chalmers claims that there is no contradiction, detectable a priori, in describing a possible world physically identical to the actual world, but lacking phenomenal experience. As we will see, however, examining the concepts currently involved in the proposition that zombies are conceivable is not sufficient to determine whether zombies actually are conceivable. Instead, we must also determine, for every possible world we are considering, whether the truth of this proposition would conflict with any other description of that world. And this may require our knowledge of further concepts that are not (now) available to us. The details of this argument

[42] This way of presenting Chalmers' conceivability argument is borrowed from Balog (forthcoming), 5.

will be explored in the next chapter. In what follows, I argue against the conceivability of zombies, and claim that not only is Chalmers guilty of equivocation, but that given his requirements on conceivability, there is reason to be skeptical about the adequacy of our conceivability judgments.

CHAPTER TWO

SUPERVENIENCE AND THE FALLACY OF EQUIVOCATION

2.1. General Remarks

Since the publication of *The Conscious Mind* (1996), a large number of objections to Chalmers' argument against physicalism have appeared in the philosophical literature. It is well beyond the scope of this thesis to present and assess every argument against Chalmers. However, a general trend among these criticisms is a focus on Chalmers' claim that zombies – that is, entities physically identical to human beings but lacking phenomenal consciousness – are conceivable, and that thus phenomenal consciousness cannot be explained by the physical facts alone. In general, two types of criticisms are made against Chalmers' account: denying his claim that the facts about phenomenal consciousness don't logically supervene on the physical facts, and denying that the conceivability of zombies entails their metaphysical possibility. This second objection is frequently accompanied by an appeal to ideas from Kripke's *Naming and Necessity* (1980), which demonstrates that the existence of necessary truths such as "water is H_2O" are only knowable a posteriori.

Concerning the first objection, it is often argued that Chalmers' position entails some form of epiphenomenalism, the view that consciousness has no effect on the physical world (e.g., Hodgson 1996, Warner 1996, Latham, 2000). Concerning the second objection, there are a number of different approaches that have been taken. Eliminative materialists, for example, have argued that facts about phenomenal conscious experience (insofar as there are such facts) are necessitated a priori by physical facts (e.g., Dennett 1996, Churchland P.M 1996, Churchland P.S 1996). And so, on this view, physically identical zombie worlds, contra Chalmers, are not conceivable. Reductionists, on the other hand, have argued that facts about phenomenal conscious experience are *not* necessitated a priori but instead a posteriori by

physical facts (e.g., Hardcastle 1996, Clark 1995, Hill and McLaughlin 1999, Loar 1999). On this view, physically identical zombie worlds are conceivable, but not metaphysically possible. And yet others have argued that while zombies themselves may be logically (or conceptually) possible, the idea of a zombie *world* is not (e.g., Yablo, 1999).

Given the vast array of literature already existing criticizing these aspects of Chalmers, in what follows, I turn to a different set of criticisms made by Mark Rowlands (2001). Rowlands attacks Chalmers' claim that consciousness does not logically supervene on the physical by arguing that Chalmers is guilty of equivocation. More specifically, he argues that Chalmers both fails to adequately distinguish between ontological and epistemological versions of supervenience, and makes fundamental errors in his conception of supervenience generally. If this is the case, which I argue it is, then Chalmers' conceivability argument against physicalism does not even get off the ground. To the best of my knowledge, Rowlands' is the only approach that attempts to undermine Chalmers' position in this way. Moreover, though it is one of the strongest arguments against Chalmers' account, it has had less attention in the literature than it deserves. This thesis aims to rectify this situation somewhat.[43]

2.2 The Problem with Chalmers' Account of Supervenience

Recall Chalmers' argument against the possibility of a reductive explanation of phenomenal consciousness goes as follows:

(1) Consciousness can be reductively explained only if it is *logically* supervenient on the physical.

(2) Consciousness is *not* logically supervenient on the physical.

Therefore, (3) Consciousness is *not* reductively explainable.

In chapter one, I claimed that the plausibility of this argument rests on an ambiguous interpretation of the concept of logical supervenience. Here, I will argue that this ambiguity derives from Chalmers' misunderstanding of the concept of supervenience in general. Chalmers provides two accounts of supervenience. On one, he characterizes it as an ontological

[43] While I find Rowlands criticism to be a serious problem for Chalmers, I do not share all of his conclusions about the nature of consciousness. Rowlands is an externalist, and holds that the phenomenal aspects of conscious experience are aspects that exist only in the directing of experience towards non-phenomenal objects.

relation of determination; on the other as an epistemological relation whereby knowledge of the A-facts entails knowledge of the B-facts. But Chalmers continually slides between ontological and epistemological interpretations of supervenience in a manner that strongly suggests that he considers them to be equivalent. It is important to note that the persuasiveness of Chalmers' argument depends on this blurring of the distinction between his ontological and epistemological interpretations of supervenience, and that he is thus guilty of problematically equivocating between the two.

Chalmers makes the same mistake with the more specific notion of logical supervenience. That is, he puts forward ontological and epistemological versions of logical supervenience, and yet treats them as equivalent. I argue that these interpretations are not equivalent, and that the plausibility of Chalmers' argument outlined above rests on understanding logical supervenience in premise (1) ontologically, but in premise (2) epistemologically. The argument is thus problematic: if we adopt his ontological interpretation of logical supervenience, premise (2) is false, and if we adopt his epistemological interpretation, premise (1) is without foundation.

Furthermore, as we have already seen, Chalmers distinguishes between logical supervenience and mere natural supervenience through the use of his "what God has to do" metaphor.[44] Using this metaphor, we get the explanation that B-properties supervene logically on A-properties if God's creation of a world with certain A-facts automatically entails the B-facts. As Chalmers puts it, where there is logical supervenience "there is a sense in which once the A-facts are given, the B-facts are a free lunch."[45] In contrast, the ontological dependence of facts on each other where there is mere natural supervenience is not so simple. Here, God's creation of a world with certain A-facts is not by itself enough to establish the B-facts. If B-properties supervene only naturally on A-properties, after fixing the A-facts, God still has more work to do in order to fix the B-facts: She has to ensure that there are laws relating the B-facts

[44] A metaphor first introduced by Kripke (1972), 253-355.
[45] Chalmers (1996), 41.

and the A-facts. Once such laws are in place, the relevant A-facts will necessitate the B-facts, though not logically since there will be possible worlds in which they do not.[46]

But Chalmers developing the idea of natural supervenience this way makes the concept of natural supervenience incoherent. This is a problem for him not only because the distinction between logical and natural supervenience is essential to his argument against physicalism, but also because his positive account maintains that consciousness is naturally supervenient on the physical.[47] And so, although Chalmers' distinction between logical supervenience and natural supervenience, properly understood, could be a legitimate one, it is not possible to accurately capture the distinction in terms of his "what God has to do" metaphor. Instead, following Rowlands I will argue that adopting the God metaphor requires abandoning the idea of supervenience in general.[48]

To reiterate, Chalmers distinguishes between logical and mere natural supervenience. In addition, he adopts two distinct conceptions of supervenience: as an ontological relation of determination, and as an epistemological relation whereby knowledge of the A-facts entails knowledge of the B-facts. Hence, there are four types of supervenience possible: logical and natural ontological supervenience, and logical and natural epistemological supervenience. Section 2.3 and 2.4 will be concerned with supervenience understood as an ontological relation of determination. I will argue that, if understood in this way, Chalmers' versions of both natural and logical supervenience are incoherent. Section 2.5 and 2.8 will be concerned with supervenience understood epistemologically. I will argue that, if understood in this way, Chalmers' versions of both natural and logical supervenience are misguided.

2.3 The Incoherence of Chalmers' Version of Natural Supervenience

As we have seen, Chalmers distinguishes between logical and natural supervenience by arguing that in the case of logical supervenience, supervenient facts (i.e., facts about B-properties) are

[46] In general, if B-properties are merely naturally supervenient on A-properties in a given world, then there *could* have been a world in which the A-facts held without the B-facts. According to Chalmers' God metaphor, the only way that the A-facts could necessitate the B-facts is if God were to create a lawful connection between them. If no such laws are created, then it follows that the B-facts would not exist. As Chalmers notes: "the B-facts are something over and above the A-facts, and their satisfaction implies that there is something new in the world" (1996, 41). Moreover, in creating the B-facts, God could implement different supervenience relations, which would result in a different distribution of B-facts. This idea will be explored in detail in the next section.

[47] Chalmers (1996), 38 and 108.

[48] Rowlands, 31.

guaranteed by the facts on which they supervene alone, but in the case of natural supervenience, they are not. Rather, in the case of natural supervenience, supervenient facts are essentially determined by the addition of connecting bridge principles (i.e., laws that link A-facts and B-facts). And so, in the case of natural supervenience, but not logical supervenience, God must not only make the distribution of subvenient facts, that is the facts about A-properties, the case, but also bring into existence the distribution of connecting bridge principles in order to fix the distribution of supervenient facts.[49] This way of thinking about supervenience relations, however, makes the concept of supervenience unintelligible.

According to Rowlands, the primary reason for introducing a concept of supervenience (whether logical, natural, or some other kind) is to express claims of the kind that the distribution of A-facts in a given world alone determines the distribution of B-facts in that world.[50] On Chalmers' understanding of natural supervenience, however, this is not the relation between A-facts and B-facts when supervenience holds. This is because on Chalmers' construal of natural supervenience, it is possible for God to create a world, W*, in which the distribution of A-facts is identical to that in W, but where the distribution of B-facts in W* is different from the distribution of B-facts in W. Let us explore this in more detail.

On Chalmers' view of natural supervenience, it is not enough that God creates the A-facts: She must do something further in order for the B-facts to come to be. In order to create the B-facts, God must create bridge principles that link the A-facts and the B-facts. And this means that "what determines the distribution of B-facts in [W and W*], is not the distribution of A-facts alone, but, also, the supervenience relations instantiated in each world."[51] And so, since the A-facts do not themselves automatically necessitate the B-facts in cases of natural supervenience, it is possible for God to create W and W*with the same distribution of A-facts, but with different supervenience relations. Because the supervenience relations instantiated in

[49] Chalmers uses the God metaphor to explain his understanding of the difference between logical and natural supervenience. This metaphor, however, is only an explanatory device. Chalmers could have explained the concept of natural supervenience without the notion of God. If we have a universe that includes not only physical facts, but also physical laws (or laws of nature), then it should not matter where those laws came from or how they were established. Indeed, Chalmers' distinction between logical and natural supervenience does not depend on the existence of God.

[50] Rowlands, 30.

[51] Rowlands, 30.

W and W* are different, it follows, on Chalmers' view, that the distribution of B-facts in these worlds will also be different.

Following Rowlands (2001), let us call the supervenience relations that obtain between A-facts and B-facts SR1 relations. As we have seen, on Chalmers' account, SR1 relations are not solely comprised of A-facts; rather, they are also comprised partly, but essentially, of B-facts. Thus, SR1 relations can be thought of as A-B facts. And so, on Chalmers' understanding of natural supervenience, what determines the distribution of the B-facts is both the distribution of the A-facts and the distribution of SR1 relations. However, thinking about natural supervenience in this way is inconsistent with the normal concept of supervenience, since, properly understood, supervenience is the idea that in a given world, the distribution of A-facts *alone* determines the distribution of B-facts in that world. It seems, therefore, that Chalmers has an illegitimate understanding of supervenience. As Rowlands points out, "if [B]-facts naturally supervene on [A]-facts, *as Chalmers understands natural supervenience*, then [B]-facts do not seem to supervene on [A]-facts at all."[52] Unfortunately, Chalmers' misconstrual of the concept of supervenience is not confined to his account of natural supervenience; rather, he builds on his misunderstanding in a way that also affects his account of logical supervenience.

2.4 The Incoherence of Chalmers' Version of Logical Supervenience

As we have seen, on Chalmers' view of natural supervenience, bridge principles, or laws, connecting A-facts and B-facts are further facts about the world. This, however, is not the case with logical supervenience. According to Chalmers, in the case of logical supervenience, "bridge laws are not further facts about the world; rather, [they are] themselves supervenient on the low-level [or subvenient] facts."[53] Thus, in the case of logical supervenience, but not natural supervenience, the distribution of A-facts in a given world alone determines not only the B-facts but also the distribution of supervenience relations in that world. In other words, given the way the A-facts are distributed, there can only be one distribution of logical supervenience relations. And so, in the case of logical supervenience, God creating W and W* with the same distribution of A-facts but with distinct supervenience relations is not a genuine possibility. This is a serious problem for Chalmers, since his argument against a reductive

[52] Rowlands, 30-31.
[53] Chalmers (1996), 108.

explanation of phenomenal consciousness rests on the idea that it is logically possible to have a world that is physically identical to ours (i.e., the distribution of A-facts is identical), but which lacks phenomenal consciousness (i.e., the distribution of B-facts is distinct).[54]

According to Rowlands, Chalmers' view of supervenience relations, in the case of logical supervenience, provides us with "the supervenience of supervenience relations upon [A]-facts, [or,] SR2 relations."[55] Recall that Chalmers says that in the case of logical supervenience the bridge laws are supervenient on the subvenient facts. Unlike SR1 relations, then, which hold between A-facts and B-facts, SR2 relations hold between A-facts and SR1 relations. And so, SR2 relations are A-SR1 facts. Since SR2 relations are composed partly, but essentially, of SR1 relations, and SR1 relations are not composed purely of A-facts, it follows that SR2 relations are also not composed purely of A-facts. Thus, if the distribution of B-facts is determined partly by SR2 relations, then it follows that the distribution of B-facts is, again, not determined by the A-facts themselves alone. Chalmers' understanding of logical supervenience therefore has the same problem as his natural supervenience. It should now be clear that on Chalmers' construal of supervenience, whether natural or logical, we are forced to give up the idea that B-facts are determined exclusively by A-facts, exactly when the B-facts are supposed to be supervenient on the A-facts. Adopting this view therefore abandons the fundamental nature of the concept of supervenience. Chalmers' understanding of both natural and logical supervenience is thus internally inconsistent.

2.5 Reified Vs. Non-reified Interpretations of Supervenience Relations

Chalmers' misunderstanding of the concept of supervenience is due to a conflation of two different interpretations of natural supervenience. Properly understood, natural supervenience is the idea that within a given set of naturally possible worlds, the A-facts necessitate the B-facts. That is, "within class of worlds $W_{1,}$ where set $S = \{s_1 \ldots s_2\}$ of supervenience relations obtain, A-facts determine B-facts."[56] While there is nothing wrong with the concept of natural supervenience understood in this way, it must not be confused with the version endorsed by Chalmers: "within class of worlds $W_{1,}$ where set $S = \{s_1 \ldots s_2\}$ of supervenience relations obtain,

[54] Chalmers (1996), 123.
[55] Rowlands, 31.
[56] Rowlands, 32.

A-facts plus $s_1 \ldots s_2$ determine B-facts."[57] On Chalmers version of natural supervenience, what determines the distribution of supervenient facts is a combination of subvenient *and* supervenient facts. As we have seen however, this is not an accurate understanding of the concept of supervenience. Chalmers misconstrues the concept of natural supervenience because he has an even deeper misunderstanding of the nature of supervenience relations.

Chalmers' mistake is his assumption "that supervenience relations form the *basis* of the dependency between subvenient and supervenient facts, when, in fact, they are simply *reflections* of a dependency that is fixed by other means."[58] To help explain Chalmers' mistake, Rowlands distinguishes between *reified* and *non-reified* interpretations of supervenience relations. On a *non-reified* interpretation, supervenience relations involve only two sets of properties or facts being related. This view entails that if something more than the A-facts is needed for the B-facts to be the case, then we do not have supervenience of the B-facts on the A-facts. It is important to recognize that "the commitments of the non-reified construal include nothing more than [the A-facts] and [the B-facts]."[59] That is, the supervenience relations that hold between the A-facts and the B-facts have no role in fixing the distribution of B-facts. Instead, the distribution of B-facts is determined by the nature and distribution of A-facts. On a non-reified interpretation, therefore, talking about supervenience relations is nothing more than talking about the relatedness of A-facts and B-facts, and recognizing that this relatedness is a dependency that cannot be explained by the supervenience relations that hold between them. The non-reified view understands supervenience relations to be constituted by the dependency of the supervenient facts on the subvenient, not a factor in an explanation of this dependency.

On a *reified* interpretation, on the other hand, the relatedness of A-facts and B-facts *is* explained in terms of the supervenience relations that hold between them. On this view, the distribution of B-facts in a given world must be explained in terms of the nature and distribution of A-facts in that world, plus the supervenience relations instantiated in that world. According to Rowlands, "reference to supervenience relations on [a] reified interpretation, is

[57] Rowlands, 32.
[58] Rowlands, 32.
[59] Rowlands, 32.

reference to a factor that, in part, fixes the relation between [A]-facts and [B]-facts."[60] In other words, talking about supervenience relations is not simply talking about the relatedness of A-facts and B-facts; rather, it includes talking about how this relatedness comes to be determined. On a reified interpretation, therefore, supervenience relations themselves form the basis of the dependency between A-facts and B-facts.[61]

Given Chalmers' account of both natural and logical supervenience, it is clear that he adopts a reified interpretation of supervenience relations. However, on a reified interpretation, supervenience relations are considered the source of the dependency between subvenient facts and supervenient facts. And, as we have seen, thinking about supervenient relations in this way makes the concept of supervenience (whether natural or logical) inconsistent. "Once we think of the supervenience relation as the partial basis of the dependency in this sort of way, we are forced, in effect, to abandon the claim that [B]-facts supervene on [A]-facts."[62] In order to avoid rendering the concept of supervenience internally inconsistent, therefore, a non-reified interpretation of supervenience relations must be adopted.

As we have seen, on a non-reified interpretation, supervenience relations are not responsible, in whole or in part, for the dependency between subvenient and supervenient facts. Instead, the dependency is determined by the nature and distribution of subvenient facts. When we assert that there is a supervenience relation that holds between A-facts and B-facts, we are simply asserting that "there is a certain type of correlation - roughly, asymmetrical dependence - between [A-facts] and [B-facts], a correlation that is explained by the nature of [A-facts]."[63] This is not to say, however, that there is a *relation* holding between A-facts and B-

[60] Rowlands, 33.
[61] Rowlands notes that his distinction between reified and non-reified interpretations of supervenience relations is similar to a distinction that is often made in the area of causal determination: a distinction between dispositional properties and the underlying structural bases of those properties. Rowlands claims that similar to supervenience relations, "to talk of an object's dispositional properties is not to talk of the *basis* of that object's behavior; rather, such talk is a reflection or *expression* of a basis that is fixed by other means, namely, by the underlying structural properties." ((2001), 33). While I do not think that supervenience relations are a species of causal determination (as does, for example, Searle (1992)), I do agree with Rowlands that there seems to be a parallel distinction between what fixes the relation between A-facts and B-facts, and the relation itself. And, it is recognizing this distinction that is essential to understanding why Chalmers' argument makes the concept of supervenience unintelligible.
[62] Rowlands, 34.
[63] Rowlands, 34.

facts which provides the basis of the asymmetrical dependence of B-facts on A-facts. On a non-reified interpretation, therefore, supervenience relations are irrelevant in fixing the distribution of B-facts.

There is nothing inconsistent about the concept of natural supervenience when it is understood in non-reified terms. According to Rowlands, what introduces the inconsistency into Chalmers' account of natural supervenience is "his reified gloss of [the] concept, a gloss surreptitiously introduced via his use of the "what God has to do" metaphor."[64] Properly understood, natural supervenience and logical supervenience are two distinct concepts; however, as I have tried to show, if we understand this distinction in terms of Chalmers' God metaphor, it collapses. It is a misunderstanding of the concept of supervenience to argue, as Chalmers does, that in cases of logical supervenience, but not natural supervenience, once the subvenient facts are established, the supervenient facts are automatically guaranteed. For this is also true in cases of natural supervenience. The dependence of supervenient on subvenient facts alone holds for genuine supervenience of any kind.[65] In the case of natural supervenience, therefore, God has only to create the subvenient facts in order to fix the distribution of supervenient facts.

Understanding the difference between logical and natural supervenience in terms of the "what God has to do metaphor" is essential to Chalmers' argument against a reductive explanation of phenomenal consciousness. It is this understanding of logical supervenience that forms the basis for premise (2) of his argument. However, if we understand logical supervenience in this way, then the distinction between logical and natural supervenience collapses. To argue, as Chalmers does, that it is only true in cases of logical supervenience that the A-facts alone necessitate the B-facts depends on accepting an illegitimate understanding of supervenience, for, as we have seen, this is also true in cases of natural supervenience. And so, "[if] we understand logical supervenience in the way Chalmers suggests, that is, in terms of the idea that supervenient facts come for free, then a*ll supervenience is logical supervenience*."[66]

[64] Rowlands, 34.
[65] As Rowlands argues, "in any genuine case of supervenience, once we fix the subvenient facts, the supervenient facts are a free lunch – always." (p. 34).
[66] Rowlands, 34.

As we have seen, Chalmers argues that consciousness naturally supervenes on the physical. However, on his ontological understanding of the notion of supervenience, the distinction between logical and natural supervenience collapses. And so, Chalmers must (on pain of contradiction) agree that consciousness logically supervenes on the physical. If we adopt Chalmers' construal of ontological supervenience, therefore, we must reject premise (2) of his argument - the claim that consciousness is not logically supervenient on the physical.

Thus far, I have tried to show that when supervenience is understood as an ontological relation of determination, Chalmers' versions of both logical and natural supervenience are incoherent. I will now explore Chalmers' epistemological versions of logical and natural supervenience. In what follows, I argue that if we understand logical supervenience in this way, there is no reason for believing premise (1) of Chalmers' argument against physicalism.

2.6 Logical Supervenience and Natural Supervenience as Epistemological Concepts

Recall that on Chalmers' epistemological interpretation, B-facts supervene on A-facts when knowledge of the A-facts is sufficient for knowledge of the B-facts. Chalmers uses Laplace's Demon (henceforth LD) - a hypothetical super-being capable of knowing an infinity of facts and making inferences from them - to explain that when supervenience holds, the only necessary information for knowledge of the supervenient facts is knowledge of the subvenient facts.[67] The idea is that, in the case of *logical supervenience*, if LD was in possession of all of the subvenient facts of the universe, she would be able to "read off" from these all of the supervenient facts of the universe.[68]

In the case of Chalmers' version of natural supervenience, however, LD's capacity for reading off the supervenient facts from knowledge of the subvenient facts is not guaranteed. This is because, as we have seen, in the case of natural supervenience, in addition to the subvenient facts there are bridge principles or laws that must hold between the A-facts and the B-facts. Therefore, in order to read off the supervenient facts from the subvenient facts, LD

[67] Chalmers' example is derived from a thought experiment originally introduced in 1814 by Pierre-Simon Laplace. Laplace's thought experiment involves a hypothetical entity, envisioned such that if it knew the precise location and movement of every atom in the universe, it could then reveal the entire course of cosmic events, past, present, and future. It should be noted, however, that Laplace referred to this entity as a "hypothetical intellect." The idea of "Laplace's Demon" was later adopted by philosophers defending the existence of a free will. For further discussion, see Laplace (2007), 4.

[68] Chalmers (1996), 35.

would have to know both the subvenient facts, and also the supervenience relations that hold between them and the supervenient facts. And, as we have seen, given Chalmers' reified understanding of supervenience relations, this would require knowledge of both the subvenient facts *and* the supervenient facts. In other words, to "read off" the supervenient facts from the subvenient facts, LD must already possess (at least) the concepts that represent the supervenient facts. Thus, when natural supervenience holds, if LD is only in possession of the concepts that represent the A-facts, then she is committed to denying not only the existence of the laws that link the A-facts and the B-facts, but she must also deny that B-facts are essentially composed of A-facts. As we can see, however, doing this would be the result of LD's being conceptually confused.

Chalmers' epistemological understanding of the distinction between logical and natural supervenience differs from a preferable, and more common, epistemological interpretation. To understand the distinction between logical and natural supervenience on this more common interpretation, imagine the following scenario: suppose that we are presented with two objects, x and y, that are indiscernible with respect to a subvenient property, A, and are asked to consider their possession of a supervenient property, B.[69]

On one hand, if B-properties *logically supervene* on A-properties, then we could not deny that objects x and y are also indiscernible with respect to their B-properties. In cases of logical supervenience, denying this indiscernibility would involve a conceptual confusion. As Rowlands notes, "any two objects which instantiate the same [A]-facts also, necessarily, instantiate the same [B]-facts. [We] know [we] must believe this on pain of being conceptually confused or logically inconsistent."[70]

On the other hand, if B-properties *naturally supervene* on A-properties, then we could claim that objects x and y, while indiscernible with respect to their A-properties, are discernible with respect to their B-properties. In the case of natural supervenience, denying this indiscernibility would not necessarily involve a conceptual confusion. Instead, according to Rowlands, "[it] would commit us to denying a relevant natural law, and [although] this might be a manifestation of extreme [empirical] ignorance on our part, [it] is an entirely different matter

[69] Rowlands, 37-38.
[70] Rowlands, 41.

[than] exhibiting conceptual confusion."[71] Thus, when natural supervenience holds, although the distribution of A-facts determines the distribution of B-facts, we can assert that objects x and y, while indiscernible with respect to their A-properties, are discernible with respect to their B-properties without, necessarily, being conceptually confused.

On an epistemological interpretation, the distinction between logical supervenience and natural supervenience does not amount to a difference in the way that the B-facts are determined; rather, it corresponds to a difference in our knowledge of the way in which the B-facts are determined. As we will see, however, the problem with Chalmers' epistemological interpretation of this distinction is his idea of "reading off" the supervenient B-facts from the subvenient A-facts. As I will try to show, LD's capacity for "reading off" the B-facts from the A-facts is irrelevant to the issue of whether the B-facts logically supervene on the A-facts. Instead, this issue turns on whether it is possible to claim that two objects indiscernible with respect to their A-facts, are distinct with respect to their B-facts. Moreover, I argue that in the case of logical supervenience, for LD to "read off" the supervenient B-facts she must possess the relevant B-concepts.

2.7 Logical Supervenience and the Idea of "Reading off" Facts

As an example of the problem with Chalmers' epistemological understanding of logical supervenience, consider his treatment of moral supervenience (i.e., the supervenience of moral facts on natural facts, where natural may include mental or physical properties).[72] Moral supervenience is generally thought to be an instance of logical supervenience (e.g., Blackburn, 1971, 1985 and Hare, 1984). However, if logical supervenience is understood, as Chalmers understands it, in terms of the ability to "read off" supervenient facts from subvenient facts, then it *fails* to be an instance of logical supervenience. For Chalmers, the claim advanced by Moore (1922), that there is no conceptual connection between natural facts and moral facts, is relevant to the question of whether moral facts logically supervene on natural ones.[73] However, if there is nothing about the meaning of a term such as "goodness" that allows that facts about

[71] Rowlands, 38.
[72] Chalmers (1996), 83.
[73] Rowlands, 38.

"goodness" should be entailed by natural facts, then it seems that "goodness" cannot be logically supervenient on natural facts.

But determining whether natural facts entail moral facts, or whether moral facts can be "read off" from natural facts, is not necessarily relevant in determining whether moral facts supervene logically on natural facts. As Rowlands notes, "moral facts are more commonly thought to supervene logically on natural [facts], but this has nothing to do with the possibility of "reading off" the former from the latter, and nothing to do with the latter somehow entailing the former."[74] Instead, the claim that moral facts supervene logically on natural facts, derives from the idea that to assign different moral properties to objects that are indiscernible with respect to their natural properties would involve a conceptual confusion. Let us explore this idea further.

Suppose for example that we have, for moral evaluation, two persons, Ghandi, and his twin earth counterpart Fhandi. Now, suppose that Ghandi and Fhandi share any natural properties that might conceivably be relevant to their moral evaluation. For example, both Ghandi and Fhandi perform the same actions (e.g., they led nationwide campaigns to ease poverty and expand women's rights, undertook long fasts, defended the rights of immigrants, etc.). Moreover, both men perform these actions with the same intentions or motives. Essentially, whatever natural qualities you think are relevant to the moral evaluation of Ghandi, assume that these qualities are also shared by Fhandi. Given this information, it would not make sense to assign different moral evaluations to Ghandi and Fhandi. Thus, if Ghandi is morally praiseworthy, then Fhandi must be as well. According to Rowlands, "it would be logically incoherent (in the broad sense), or conceptually confused, to assign a different moral evaluation to each person."[75]

It is important to recognize that there is nothing in this example that requires the possibility of our "reading off" the moral facts from the natural facts. That is, being presented with Ghandi's natural qualities does not automatically allow us to determine his moral qualities. Rather, as Rowlands notes, being presented with Ghandi's natural qualities simply tells [us] that, "*on pain of contradicting [ourselves]* (in a broadly logical sense), [we] cannot assign

[74] Rowlands, 39.
[75] Rowlands, 39.

[Ghandi] a certain moral evaluation while assigning another person, indiscernible in all relevant natural respects, a distinct one."[76]

Even in the case of global logical supervenience, knowledge of the natural facts is not sufficient to determine the distribution of moral facts in a given world. In fact, knowledge of the natural facts is not sufficient to determine whether moral facts even exist in a given world. "All logical supervenience tells [us] is that if there is another world indiscernible (in a non-epistemic sense) from the world under consideration, then this other world will possess the same distribution of moral facts."[77] In other words, in the case of global logical supervenience of the moral on the natural, if two worlds are indiscernible with respect to their natural properties, they must also be indiscernible with respect to their moral properties. To claim otherwise would involve a conceptual confusion.

Thus far, we have seen that on an epistemological interpretation, the difference between logical supervenience and natural supervenience amounts to whether it is possible to coherently claim that two objects indiscernible with respect to the instantiation of their subvenient facts, are distinct with respect to the instantiation of their supervenient facts. Consider an example frequently referred to by Chalmers: the logical supervenience of biological facts on physical facts. We can imagine, Chalmers tells us, that LD, who knows the location of every particle in the universe, would be able to straightforwardly "read off" the biological facts once given all of the physical facts. [78] On Chalmers' view, the microphysical facts are sufficient for LD to construct a model of the microscopic structure and dynamics of the world, which, in turn, is sufficient information for her to straightforwardly deduce the macroscopic biological structures and dynamics. However, as we have seen, LD's capacity for "reading off" the biological facts from the physical facts is irrelevant to the issue of whether the biological facts logically supervene on physical facts.

What is relevant is whether claiming that two objects indiscernible with respect to their physical properties are distinct with respect to their biological properties involves a conceptual confusion. In the case of natural supervenience, however, it would be possible to make this

[76] Rowlands, 39-40.
[77] Rowlands, 40.
[78] Chalmers (1996), 35.

claim without, necessarily, exhibiting conceptual confusion. Instead, to make this claim would be to deny some relevant natural law(s), and this would be the result of empirical ignorance, not conceptual confusion. And so, contrary to Chalmers' understanding, when logical supervenience is understood in an epistemological sense, it *is* possible that macrophysical facts (e.g., biological facts) fail to supervene logically on microphysical facts. According to Rowlands:

> If [a person] can (i) assert that two objects have the same microphysical properties, and (ii) deny that they share the same macrophysical properties, and, in so doing, (iii) exhibit only gross empirical ignorance rather than conceptual confusion, then this is sufficient to show that macrophysical properties do not logically supervene on microphysical ones, where, to reiterate, this supervenience is understood epistemologically.[79]

In other words, what an epistemological understanding of the logical supervenience of phenomenal on physical facts amounts to is this: any two objects which instantiate the same A-facts (physical facts) also, necessarily, instantiate the same B-facts (phenomenal facts). And so, if objects x and y instantiate the same A-facts, we know that they must also instantiate the same B-facts. To deny this would involve being conceptually confused or logically inconsistent. It is important to recognize, however, that logical supervenience does not, by itself, allow us to "read off" that x and y instantiate certain B-facts from knowledge of what A-facts they instantiate.

2.8 LD's Inability to "Read off" Facts

As we have seen, the concept of logical supervenience is not typically understood as Chalmers conceives it. That is, logical supervenience is not normally understood as a relation that entails the ability to "read off" supervenient facts from knowledge of subvenient facts. This is not to say, however, that it cannot be understood in this way. Take, for example, Jaegwon Kim's (1984) argument concerning the existence of correlations between supervenient properties and subvenient properties.

According to Kim, the (strong) supervenience of B-properties on A-properties entails the existence of (i) modal bi-conditionals connecting B-properties with disjunctive A-properties, and (ii) modal conditionals running from A-properties to B-properties.[80] On this view, supervenient

[79] Chalmers (1996), 35.
[80] Kim, 169-170.

properties necessarily have coextensive properties in the set of properties on which they supervene. This, then, is an entailment of logical supervenience. Understood in this way, it might be thought that logical supervenience does entail the possibility of "reading off" the supervenient facts from the subvenient facts. That is, logical supervenience might be thought to entail bridge principles that connect A-properties and B-properties, and these principles allow the possibility of "reading off" B-properties or facts, from A-properties or facts. However, as Rowlands points out, "the conclusion of Kim's argument is an ontological one, pertaining to the sorts of relations that exist in the world, relations entailed by the strong supervenience of one set of properties on another."[81] Indeed, Kim himself acknowledges that his conclusion is not meant to be taken as an epistemological claim.[82] The issue then, is whether an epistemological construal of supervenience, in terms of "reading off" supervenient from subvenient facts, is obtainable from the relation of logical supervenience.

As we have seen, LD, who possesses knowledge of all of the basic facts of the universe, is, on Chalmers' view, able to "read off" all facts that are *logically supervenient* on the basic facts. This means, therefore, that LD's ability to "read off" a given set, S, of facts depends on two conditions obtaining: (i) LD has knowledge of all of the relevant basic facts, and (ii) all facts that are members of S are logically supervenient on these facts. However, to claim that LD's capacity for "reading off" the logically supervenient facts depends only on her knowledge of the basic facts is a deeply misguided idea. To see this, let us first consider what Chalmers means by "basic facts."

Here, it is important to recognize the distinction between strong and weak supervenience. B-properties supervene *strongly* on A-properties, if, for any two possible worlds, w1 and w2, and for any two individuals, x in w1 and y in w2, if x and y are identical concerning A-properties, they are also identical concerning B-properties. For example, if psychological properties strongly supervene on physical properties, then any two physically identical people will also be psychologically identical. In comparison, B-properties supervene only *weakly* on A-properties if, for any two individuals, x and y, in the *actual world*, if x and y are identical concerning A-properties, they are also identical concerning B-properties. If psychological properties only weakly supervene on physical properties then two physically identical people in the *actual world* will also be psychologically identical. The supervenience relation (if it is weak) only holds in the actual world, and can fail to hold in other possible worlds. The difference between strong and weak supervenience, therefore, depends on whether the correlation between supervenient properties (i.e., B-properties) and base properties (i.e., A-properties) hold within the actual world only, or across all possible worlds.

[81] Rowlands, 41.

[82] Kim, 173.

A basic fact is the instantiation of a basic property, where "basic property" refers to "the fundamental properties that are invoked by a completed theory of physics."[83] Further, when Chalmers talks about the supervenience of properties on the physical, he intends that "physical property" be understood as "basic property." In this sense, physical properties might include such things as mass, charge, and spatiotemporal position. However, "such high-level properties as juiciness, lumpiness, giraffehood, and the like are excluded, even though there is a sense in which these properties are physical."[84] Therefore, when a set of facts supervene on the physical, they supervene on the fundamental properties invoked by a completed physics, and not physical properties in a broader sense. And so, the basic facts of which LD has exhaustive knowledge are the physical facts in this restricted sense.

Suppose then, for example, that LD inhabits a world where a set of non-basic facts, S, logically supervenes on basic facts, where basic facts are physical facts in Chalmers' restricted sense. If LD is able to "read off" the members of S equipped only with knowledge of the basic, physical, facts, then she must, according to Rowlands, "be able to somehow construct knowledge of the non-basic facts from [her] knowledge of the basic ones."[85] The question is: how can she do this?

It is apparent that we cannot allow for LD to possess knowledge of the relations entailed by strong supervenience. This is because these are "the modal conditionals and bi-conditionals that Kim argues are consequences of (strong) supervenience."[86] So, to allow that LD has knowledge of these relations would make the idea of her "reading off" non-basic facts from basic facts vacuous. The relations that are entailed by strong supervenience are, in effect, composite facts, necessarily comprised of both basic and non-basic facts. And so, if LD is to have knowledge of these composite facts, she must be able to construct it from her knowledge of the basic facts alone. Once again, however, we are left with the question: how can she do this?

[83] Chalmers (1996), 33.
[84] Chalmers (1996), 33.
[85] Rowlands, 42.
[86] Rowlands, 42.

According to Rowlands, "the reason Chalmers thinks LD could construct such knowledge is that he thinks all logically supervenient facts are, essentially, instantiations of *causal* or *functional* properties."[87] Understood this way, the supervenience relations that hold between the basic and non-basic facts are themselves logically supervenient on the basic facts. When logical supervenience holds between basic facts and non-basic facts, therefore, it is possible for LD to "read off" the supervenience relations from the basic facts. And so, for Chalmers, "the ability of LD stems ultimately from the nature of logically supervenient [or non-basic] facts."[88]

On Chalmers' view, non-basic logically supervenient facts are instantiations of functional properties, where a functional property can be analyzed in terms of its place in a nexus of causal relations. Thus, if a basic property, or a combination of such properties, fits into a place in the causal nexus, then the basic facts, the instantiation of these basic properties, explain the non-basic facts. Let us suppose Chalmers is correct that non-basic properties can be analyzed in terms of causal roles. Take, for example, a table. There are basic facts about the table, including (among other things) its composition, position, and distribution of properties. The solidity of the table, however, is a non-basic fact, and it logically supervenes on the basic facts. Does it follow that LD is able to "read off" the table's non-basic facts from her knowledge of the basic facts? If LD possesses knowledge of the causal analyses of the non-basic properties of the table, then this certainly seems to be the case. However, there is no reason to think that these causal analyses should be taken to be basic facts. And so, if Chalmers is serious about the idea of logical supervenience entailing that LD can "read off" non-basic facts from her knowledge of basic facts, then LD must be able to construct the causal analyses of non-basic properties from her knowledge of the basic facts. Once again, however, we return to the same question: how can she do this?

The answer, of course, is that LD cannot do this. Any attempt LD makes to "read off" the supervenient facts from the subvenient ones is going to be blocked by the *holism of the physical.* As we have seen, logically supervenient facts are the instantiation of logically supervenient properties, and so, "if we adopt the sort of causal-functional analysis of such properties endorsed by Chalmers, then we are led, pretty straightforwardly, into a holistic

[87] Rowlands, 42.
[88] Rowlands, 43.

account of their identity conditions."[89] For example, the instantiation of a logically supervenient property consists in the filling of a causal role. However, "the properties referred to in the specification of [a causal role] will themselves be logically supervenient on more basic properties, and so, their instantiation will, in turn, consist in the filling of further causal roles, and so on."[90]

To help understand why the holism of the physical poses a problem for LD, it may be beneficial to look at an example provided by Rowlands (2001) concerning the *holism of the mental.* Imagine that LD has a cousin, Watson's Demon (henceforth WD), who, according to demonic legend, is able to "read off" mental facts from her knowledge of behavioral facts. It is argued, however, that because of the holism of the mental, there could never be a demon with this ability. According to Rowlands, "the problem the holism of the mental provides for WD is that any attempt to specify in behavioral terms what a given mental state is, eventually but inevitably, will make reference to some other mental state."[91] This is because a given mental state will have consequences for behavior only in combination with other mental states. Take, for example, the interaction between beliefs and desires: running from a bear likely involves the belief that the bear is going to attack you, but why should you run from the bear unless you also desired not to be attacked? Thus, WD cannot "read off" mental facts equipped only with her knowledge of behavioral facts. Instead, to "read off" a given mental fact requires knowledge of both behavioral facts and further mental facts. Rowlands claims that this is because "mental states form a complex, holistic, network, and the identity conditions of any given mental state depend, at least in part, on their position in this network."[92]

In the same way that WD's attempt to "read off" mental facts from behavioral facts is blocked by the holism of the mental, LD's attempt to "read off" supervenient facts from subvenient facts is blocked by the holism of the physical. As we have seen, given a causal-functional analysis, non-basic facts are largely interrelated, and therefore, cannot be "read off"

[89] Rowlands, 44.
[90] Rowlands, 44.
[91] Rowlands, 44.
[92] Rowlands, 44.

from knowledge of the basic facts alone, not even in a world where all non-basic facts logically supervene on basic facts. Let us explore this in more detail.

Imagine a non-basic property, the molecular property MP, and the instantiation of that property, the molecular fact MF. Following Chalmers, suppose that MP can be analyzed in terms of a causal role, C_1. Further, suppose that there is a basic property, AP, which when instantiated as a basic fact, AF, fills causal role C_1. As we can see, the instantiation of MP consists in the instantiation of AP. Moreover, the instantiation of AP explains the instantiation of MP. From this, does it follow that LD can "read off" MF from AF?

In order for LD to "read off" MF from AF, she must know that MF is the instantiation of a property MP that is defined by a causal role C_1. As Rowlands notes, however, "that MP is defined by causal role C_1 is not knowledge of an A-fact, it is knowledge of an M-fact."[93] And so, if LD is to "read off" the M-facts equipped only with knowledge of the A-facts, she must be able to construct the knowledge that MP is defined by a causal role C_1 from her knowledge of the A-facts alone. However, due to the holism of the physical, any attempt by LD to "read off" MF from AF will fail. Rowlands explains:

> The properties that will be referred to in the specification of C_1 are M-properties, [and] hence defined by further causal roles C_2... C_n. Thus, in order to "read off" MF from AF, LD must know that MP is that property which, when instantiated in conjunction with property MP_n, causes the instantiation of MP_{n+x}… etc., etc.[94]

At any given point, the M-properties will themselves be defined in terms of further causal roles. And so, any attempt by LD to "read off" MF from AF will result in her being caught in a web of M-concepts in precisely the same way that WD was caught in her attempt to "read off" mental facts from behavioral facts.

As we can see, when non-basic facts logically supervene on basic facts, then what knowledge of the basic facts provides LD is not the deductive ability to "read off" the non-basic facts. Rather, it provides the "ability to combine knowledge of the basic facts, with the *non-basic* causal analyses of non-basic properties, to work out what instantiations of basic

[93] Rowlands, 45.
[94] Rowlands, 45.

properties the instantiation of a given non-basic property consists in."[95] Thus, Chalmers' idea that LD can "read off" the non-basic facts from knowledge of the basic facts alone is mistaken, even in a world where non-basic facts logically supervene on basic facts.

To make matters worse for Chalmers, in his explanation of the concept of logical supervenience, he actually recognizes that LD is not going to be able to "read off" logically supervenient facts from knowledge of the subvenient facts alone. Chalmers writes:

> In general, when B-properties supervene logically on A-properties, we can say that the A-facts entail the B-facts, where one fact entails another if it is logically impossible for the first to hold without the other. In such cases, Laplace's Demon could read off the B-facts from a specification of the A-facts, *as long as it possesses the B-concepts in question.*[96]

And so, Chalmers acknowledges that for LD to "read off" the logically supervenient facts from their subvenient bases, she must possess the relevant B-concepts. However, a problem emerges when we consider what a B-concept is, because "possession of B-concepts will almost certainly require possession of B-facts for the simple reason that the content of B-concepts is specified in terms of B-facts."[97] For example, the concept of "giraffehood" is, for Chalmers, a B-concept. What knowledge would LD need in order to posses this B-concept? First, she would need to know enough of the relevant characteristics of a giraffe. That is, LD would need to know something of the following sort: something is a giraffe if… (The dots are filled in by specifying, for example, the relevant details of genetic structure). However, this means that LD will have to know certain facts. She will have to know, for example, that something counts as a giraffe only if it fits a certain genetic profile, something clearly not basic, not an A-fact. And so, as Rowlands notes, "to give LD free knowledge of B-concepts is, in effect, to render vacuous the idea that LD can "read off" all facts that are logically supervenient on A-facts simply from possession of A-facts alone."[98]

To reiterate, when you have an analysis of a given B-property, then you have an account of what is included in the instantiation of that property. Once you know what A-facts fill the role specified in the analysis, then it is possible to "read off" the relevant B-facts. This is true in cases

[95] Rowlands, 46.
[96] Chalmers (1996), 36, emphasis mine.
[97] Rowlands, 46.
[98] Rowlands, 46.

of *both* logical and natural supervenience for Chalmers. In fact, the distinction between logical and natural supervenience is irrelevant to the idea of "reading off" supervenient facts from subvenient facts. This distinction could only be relevant if we were willing to allow that logical and natural supervenience denoted two ontologically distinct kinds of relation. As Rowlands notes, "this would allow us the sort of claim often made by Chalmers, that in a case of logical supervenience the supervenient facts *consist in* the subvenient ones, whereas in a case of natural supervenience they do not."[99] However, as we have seen, this claim is problematic.

To summarize, we should conclude that Chalmers' idea of "reading off" logically supervenient facts from their underlying subvenient bases is an entirely misguided one. As I have tried to show, any attempt by LD to "read off" the supervenient facts from the subvenient facts will, eventually but inevitably, require LD to be in possession of (at least some) supervenient facts. And so, as Rowlands claims, "the idea of "reading off" cannot be used to characterize the concept of logical supervenience, nor can it be used to distinguish logical from [mere] natural supervenience."[100]

2.9 Logical Supervenience and Reductive Explanation

In premise (2) of his argument, Chalmers claims that consciousness is *not* logically supervenient on the physical. The persuasiveness of this claim, however, is dependent on Chalmers' epistemological interpretation of logical supervenience, whereby it is possible to "read off" logically supervenient facts from knowledge of the subvenient facts. Chalmers argues that since it is not possible to "read off" facts about consciousness from physical facts, consciousness does not logically supervene on physical properties, and thus, cannot be reductively explained in terms of them.[101]

In this chapter, I have tried to undermine Chalmers' epistemological interpretation of logical supervenience in two ways. First, I argued that on a more common (and preferable) epistemological interpretation, to say that B-facts logically supervene on A-facts is to say that two objects indiscernible with respect to their A-properties must also be indiscernible with

[99] Rowlands, 48.
[100] Rowlands, 48.
[101] Chalmers (1996), 124.

respect to their B-properties. To claim otherwise, is to fall into a conceptual confusion. On this interpretation, the idea of "reading off" the B-facts from the A-facts is simply irrelevant. Second, I argued that Chalmers' idea of LD "reading off" logically supervenient facts equipped only with knowledge of the subvenient facts is largely misguided. In order for LD to "read off" the logically supervenient facts from the subvenient facts, she must know in what the instantiation of a supervenient property consists, and this knowledge cannot be constructed from knowledge of the subvenient facts alone.

The problem with Chalmers' argument should now be clear. On the more common epistemological interpretation of logical supervenience, most supervenient facts actually fail to logically supervene on a set of subvenient properties. As we have seen, what logical supervenience understood epistemologically means is that it is not possible to assert that two objects are indiscernible with respect to their A-properties while distinct with respect to their B-properties without involving a conceptual confusion. In this sense, however, most supervenience is not logical in character. For example, it is possible to assert that two objects are physically (in Chalmers' basic sense) identical but biologically distinct, without exhibiting conceptual confusion. Perhaps someone believes that certain biological properties, life, for example, are not physical properties of an individual, but rather, are non-physical gifts from God. Such a person might be grossly factually mistaken, but their mistake is not on account of a conceptual confusion. Indeed, "this form of biological dualism may be incorrect, but it does not seem that it is contradictory, even in the broadly logical sense."[102]

And so, if the epistemological construal of logical supervenience is understood in terms of the more common interpretation, which I have suggested it should be, then most supervenient properties will fail to supervene logically on their bases. This failure, however, does not entail the failure of reductive explanation. And so, Chalmers' claim that consciousness fails to supervene logically on the physical, does not entail the failure of a reductive explanation of consciousness. Indeed, as Rowlands notes:

> The failure of consciousness to logically supervene, epistemologically speaking, on physical facts no more entails that consciousness cannot be reductively explained, than the corresponding

[102] Rowlands, 49.

> failure of logical supervenience in the case of biological facts on (basic) physical facts entails that the former cannot be reductively explained in terms of the latter.[103]

Therefore, if we understand logical supervenience in an epistemological sense, there is no reason for believing premise (1) of Chalmers' argument (viz., that consciousness can be reductively explained only if it is *logically* supervenient on the physical). On the other hand, as we saw earlier, if logical supervenience is understood in an ontological sense, there is no reason for thinking that consciousness fails to supervene logically on the physical. And so, if logical supervenience is understood ontologically, then premise (2) of Chalmers' argument (viz., that consciousness is *not* logically supervenient on the physical) is false. If the arguments developed in this chapter are correct, then Chalmers' ontological interpretation of supervenience is incoherent, his epistemological version is misguided, and his overall argument against physicalism fails because it depends on equivocating between the two.

I will now turn to considering Chalmers' conceivability argument in a different way. According to Chalmers, the conceivability of zombies demonstrates that there is no logical entailment from physical facts to facts concerning phenomenal consciousness. As I will argue in the next chapter, however, a parallel argument can be run from the conceivability of anti-zombies - purely physical creatures that are nevertheless phenomenally conscious - to the truth of physicalism.

[103] Rowlands, 49.

CHAPTER THREE

THE INCONCEIVABILITY OF ZOMBIES

3.1 Chalmers' Conceivability Argument

Many arguments against physicalism rely on considerations of conceivability. These arguments, called conceivability arguments, have two different versions. On one, considerations of conceivability are taken to support the claim(s) that phenomenal consciousness is not identical to, realized by, or supervenient on physical properties (e.g., Kripke 1972, Nagel 1974, White 1986, Robinson 1993, Chalmers 1996, and Jackson 1998). On the other, an explanatory gap between phenomenal and physical levels of description is asserted that results in metaphysical consequences.[104] On this version, even if considerations of conceivability do not establish that phenomenal consciousness is irreducible to physical properties, we are still left with an explanatory gap between the mental and the physical – that is, we still require an explanation of the mental in terms of the physical. It is beyond the scope of this thesis to discuss the problems that are associated with both types of conceivability arguments. Thus, in this chapter I focus on the former version of conceivability arguments specifically; I am interested in Chalmers' conceivability argument invoking zombies.

According to Chalmers, zombies are beings that are physically and functionally identical to human beings, inhabiting a world which is physically identical to ours, but who lack phenomenal conscious experience.[105] In other words, zombies share all of our physical properties, but do not experience 'phenomenal feels.' In its simplest form, Chalmers' conceivability argument is as follows:

(1) Zombies are conceivable

(2) If zombies are conceivable, then zombies are metaphysically possible

[104] This argument is formulated by Levine (1998), although he does not endorse the conclusion.
[105] Chalmers (1996), 94-95.

(3) If zombies are metaphysically possible, then consciousness is non-physical

Therefore, (4) Consciousness is non-physical[106]

A more general version of this argument also often appears in the literature. It goes as follows, where P is a complete physical description of the world, including the fundamental physical laws, and Q is the existence of phenomenal consciousness (~Q, therefore, is the non-existence of phenomenal consciousness):

(1) P & ~Q is conceivable.

(2) If P & ~Q is conceivable then P & ~Q is metaphysically possible.

(3) If P & ~Q is metaphysically possible then physicalism is false

Therefore, (4) Physicalism is false.[107]

In general, physicalists respond to Chalmers' conceivability argument by either denying premise (1) - that zombies are conceivable (e.g., Dennett 1995, Shoemaker 1999, and Kirk 2005), or premise (2) - that conceivability entails metaphysical possibility (e.g., Brueckner 2001, Block and Stalnaker 1999, and Hill 1997). Following some of the literature, I will refer to the second approach as the standard objection.[108] In what follows, I provide reasons to doubt premise (1) of Chalmers' conceivability argument. My argument will be two-fold. First, I will argue that given Chalmers' requirements on conceivability, there is reason to be skeptical about the adequacy of our judgments of conceivability. More specifically, I will argue that there is good reason to be skeptical about claims of the conceivability of zombies. Second, I will turn to an argument formulated by Keith Frankish (2007). Using the same resources that Chalmers employs in his conceivability argument, Frankish constructs a parallel argument from the conceivability of anti-zombies – purely physical creatures that are nevertheless phenomenally conscious – to the truth of physicalism. Frankish argues that the anti-zombie argument can be defeated only at the cost of rendering the zombie argument itself redundant.[109] To the best of my knowledge, Frankish is only one of a handful of people to formulate an argument of this

[106] Chalmers, "Consciousness and Its Place in Nature" (2002), 249.

[107] This version of the conceivability argument appears in Balog (forthcoming), 5. Chalmers gives a similar version in "Consciousness and Its Place in Nature" (2002), 249.

[108] See for example Brueckner, 2001. The standard objection is also referred to as the "Critical Premise thesis" or CP (see Frankish (2007) and Chalmers (2007)).

[109] Chalmers' conceivability argument is frequently referred to in the literature as the "zombie argument." And so, when I am talking about the "zombie argument" here, I am referring to Chalmers' conceivability argument.

sort. This approach is starting to receive more attention in the literature (e.g., Brown 2010, Balog unpublished ms); however, it has not yet received the attention it deserves. Here, I aim to rectify this.

3.2 Chalmers' Conception of Conceivability

What does it mean to say that something is conceivable? On an epistemic understanding, conceivability depends on our cognitive abilities. That is, a situation is conceivable depending on what we know or believe, or what concepts or modes of presentation we have available or are using to think about the situation. And so, p is conceivable for subject S, if S thinks that it is possible for p to be true (e.g., Yablo 1993, p. 26), or if S does not think that p is necessarily false (e.g., van Cleve 1983, p. 37), or if S does not think that p is false (e.g., Kneale 1949, cited in Yablo 1993, p. 7), or if p (understood as a state of affairs , rather than a proposition or sentence), when judged under representation R, is thought possible (e.g., Levine 2000, p. 40). And so, on an epistemic understanding, what we can conceive depends on our power of imagination, and this, in turn, depends on the concepts we have available to us.

On an epistemic interpretation, conceivability does not entail metaphysical possibility. If what is conceivable depends on the concepts or background knowledge that we have available, then certain propositions will *seem* conceivable to us even though they are actually impossible. For example, the ancient Greeks (not realizing that Hesperus and Phosphorus are the same planet), would have found it (epistemically) conceivable that Hesperus would outlast Phosphorus. Of course, we now know that this state of affairs is impossible. What this demonstrates, however, is that knowledge of something necessary (in this case, the identity of Hesperus and Phosphorus) might be knowable a posteriori rather than a priori.[110] And so, as we can see, epistemic notions of conceivability do not overcome the standard objection, because epistemic conceivability does not entail metaphysical possibility.

Chalmers, however, has a non-epistemic understanding of conceivability. For our purpose, there are three features of his notion of conceivability that must be understood. First, to be conceivable is to be true in a conceivable world (e.g., 1996, p. 66-67). Second, the class of conceivable worlds is co-extensive with the class of logically possible worlds (e.g., 1996, p. 68).

[110] This example is derived from Worley (2003), 17. For further discussion of this point, see Levine (2000), 43-46.

Third, possibility itself is determined by conceptual coherence or incoherence.[111] Chalmers writes:

> In determining whether it is logically possible that some statement is true, the constraints are largely conceptual. The notion of a male vixen is contradictory, so a male vixen is logically impossible; the notion of a flying telephone is conceptually coherent, if a little out of the ordinary, so a flying telephone is logically possible.[112]

Here the discussion is of the logical possibility of the truth of statements, not of worlds. However, Chalmers makes clear that the logical possibility of worlds must also be determined by conceptual constraints. For instance, in describing the logical possibility of a zombie world, he writes: "there is a *logically possible* world physically identical to ours, in which the positive facts about consciousness in our world do not hold."[113] He later states: "on my usage, 'logically possible' is near enough to interchangeable with 'conceptually possible,' and is tied by stipulation to conceivability."[114] And so, on Chalmers' view, there cannot be a logically possible world which contains male vixens, since the concept of being a male and the concept of being a vixen are logically contradictory.

The question then arises: how do we extend Chalmers' understanding of the possible truth of statements to the existence of possible worlds? At first glance, the answer seems relatively straightforward: we could simply look at whether a world itself is coherent or incoherent. However, as Sara Worley points out, "the notion of conceptual coherence or incoherence applies in the first instance to statements, or descriptions, not to worlds themselves or the situations or states of affairs that are embedded in worlds."[115] And so, to say that a world is incoherent is to mean that there is some incoherence in the description of that world. Indeed, Chalmers would agree. He writes:

> Any worries about the gap between conceivability and possibility apply at the level of *statements*, not worlds: either we use a statement to misdescribe a conceived world, or we

[111] I credit Worley (2003) for pointing out the significance of these three features.
[112] Chalmers (1996), 35.
[113] Chalmers (1996), 121, emphasis mine.
Chalmers' use of "logically possible" in this claim was a source of much debate over the years (e.g., Hill and McLaughlin, 1999). Chalmers (1999) addresses the source of this misunderstanding.
[114] Chalmers (1999), 477.
[115] Worley, 18.

> claim that a statement is conceivable without conceiving of a world at all. So there seems to be no reason to deny that conceivability of a world implies possibility.[116]

Now, it is possible to have two descriptions of a single situation or state of affairs, one of which is coherent, while the other is incoherent. In these cases, do we consider the situation, and the world of which it is a part, as possible, in virtue of its coherent description, or do we consider it impossible, in virtue of its incoherent description?

If there is one incoherent description of a situation or state of affairs, that is sufficient to show that situation or state of affairs impossible. For example, we know that there are no possible worlds which contain married bachelors. This is because the concept of being a bachelor is in contradiction with the concept of being married (since by definition a bachelor is an un-married male). And so, whether there are other descriptions available of this situation which do not reveal its incoherence is irrelevant.

Furthermore, on Chalmers' view, a situation or state of affairs may be impossible even though we are not aware of a description which reveals its impossibility. Here, Chalmers has in mind examples like Goldbach's conjecture.[117] He writes:

> One might at first glance think it is conceivable that Goldbach's conjecture is false, by conceiving of a world where mathematicians announce it to be so; but if in fact Goldbach's conjecture is true, then one is *misdescribing* this world; it is really a world in which the conjecture is true and some mathematicians make a mistake.[118]

On Chalmers' view, if Goldbach's conjecture is true, then not only is its falsity impossible, it is also inconceivable. And so, the impossibility, and thus inconceivability, of its falsity, cannot depend on our being aware of a description that reveals its impossibility, since no one is as of yet aware of any such description.

As we have seen, Chalmers argues that logical possibility is to be understood in terms of conceptual coherence or incoherence. On this view, a situation is logically impossible if there exists (at least) one conceptually incoherent description of the situation. By understanding

---

[116] Chalmers (1996), 68.

[117] Chalmers (1996), 67.
For our purposes, all we need to know is that Goldbach's conjecture is a mathematical proposition, and so if true must be necessarily true. Roughly, Goldbach's conjecture is that every number greater than 2 is the sum of three primes. For further discussion see Weisstein (2010).

[118] Chalmers (1996), 67.

logical possibility in conceptual terms, Chalmers is able to avoid brute necessities. It is easy, for example, to understand why the notion of a male vixen is impossible: the concept of being a male is in contradiction with the concept of a vixen (since by definition a vixen is a female fox). The idea that there might be metaphysical necessities (e.g., propositions that are true in all possible worlds) which are not constrained by conceptual incompatibilities, however, is rather confusing. As Worley notes, "to suppose there are is to suppose that there are restrictions on the space of possible worlds above and beyond the conceptual restrictions."[119] But these restrictions would be entirely arbitrary and inexplicable.

According to Chalmers, there is no reason to believe that such a modality exists. He claims:

> It may be reasonable to countenance brute, inexplicable facts about *our* world, but the existence of such facts about the space of possible worlds would be quite bizarre. The realm of the possible (as opposed to the realm of the natural) has no room for this sort of arbitrary constraint. [120]

Indeed, if such a modality did exist, then we would have to ask why there are worlds that are conceptually possible (and thus logically possible) but metaphysically impossible (i.e., not possible in the actual world). On Chalmers' view, "it is in God's powers when creating [a] world, to do anything that is logically possible."[121] And so, if we claim that a zombie world is conceptually possible but metaphysically impossible, we must ask why God could not create such a world:

> The advocate of metaphysical necessity must say either that the possibility is coherent, but God could not have created it, or God could have created it, but it is nevertheless metaphysically impossible. The first is quite unjustified, and the second is entirely arbitrary.[122]

By understanding logical possibility in terms of conceptual coherence or incoherence, Chalmers seems able to avoid the problems presented by metaphysical necessities.

As we can see, on Chalmers' account, conceivability and metaphysical possibility come to essentially the same thing. Chalmers acknowledges this when he writes: "conceptual

[119] Worley, 19.
[120] Chalmers (1996), 137.
[121] Chalmers (1996), 138.
[122] Chalmers (1996), 138.

possibility = logical possibility = metaphysical possibility."[123] And so, as Worley notes, "to be conceivable [on Chalmers' account] is to be true in a conceivable world, but the conceivable worlds are just the possible worlds, so to be conceivable is to be true in a possible world, and, thus, to be possible."[124] In this way, the conceivability argument seems to overcome the standard objection. That is, the inference from conceivability to metaphysical possibility, that is, premise (2) of the conceivability argument, seems plausible.

I am not convinced that Chalmers' conceivability argument is able to overcome the standard objection. This is because the conceivability of zombies is not implied by the failure of a priori entailment between phenomenal properties and physical properties. Chalmers assumes too tight a connection between what is conceivable and what is possible. Discussion of this issue, however, is beyond the scope of this thesis. Instead, I want to focus on premise (1) of the argument – the claim that zombies are conceivable. Given Chalmers' restrictions on conceivability, in what follows, I argue there is reason to be skeptical about the conceivability of zombies. Let us explore this in more detail.

3.3 A Problem with Chalmers' Conception of Conceivability

As we have seen, on Chalmers' account, a state of affairs is impossible only if there is a description that reveals its incoherence (e.g., if it shown to be logically contradictory). However, we need not be aware of this description. In other words, a state of affairs may *seem* conceivable to us because the description that we are using to think about it is not sufficient to reveal its incoherence, even though there may be some other description available which does reveal its incoherence. And so, "even if conceivability is not relativized to our epistemic status, our judgments about what is conceivable must necessarily be relativized to that status."[125] Our judgments about what is conceivable, therefore, may not be reliable indicators of what *actually* is conceivable.

Chalmers essentially admits this when he argues that, on the assumption that Goldbach's conjecture is true, its falsity is not conceivable. According to Worley, "even if we grant that the truth of the conjecture precludes the conceivability of the proposition that the

[123] Chalmers (1999), 478.
[124] Worley, 19.
[125] Worley, 20.

conjecture is false, surely the most we can say is that it seems to us that it is conceivable, or, perhaps, that we simply cannot tell whether or not it is conceivable."[126]And so, we cannot rely on what *seems* to be conceivable as a guide to what *actually* is conceivable.

Furthermore, given Chalmers' understanding of conceivability, there is an additional reason to be skeptical about the adequacy of our conceivability judgments. As we have seen, on Chalmers' account, to be conceivable is to be true in a conceivable (and thus possible) world. Moreover, to determine whether a proposition is conceivable (and thus possible), we must determine whether there is a world in which the obtaining of that proposition in that world conflicts with any other description of that world. The problem, however, is that we cannot establish whether certain contradictions or inconsistencies obtain simply by examining our current concepts.

Consider again, for example, Goldbach's conjecture. As we have seen, on Chalmers' view, a state of affairs is impossible, and thus inconceivable, only if there is a conceptual incoherence. And so, there must be some incoherence in the proposition that Goldbach's conjecture is false (on the assumption that it is true). But where is this incoherence? We cannot establish that Goldbach's conjecture is true, simply by examining the conjecture itself, or the concepts involved in its formulation. If we could, then the truth of the conjecture would not still be a mathematical puzzle.

If Goldbach's conjecture is true, then its truth is presumably derivable from other mathematical truths. As Worley writes:

> There is no conceivable world in which Goldbach's conjecture is false, not because the proposition that it might be false is inherently self-contradictory or incoherent, but because these other mathematical truths, which [presumably] entail Goldbach's conjecture, are themselves true in every world.[127]

In other words, it is a logical contradiction to claim both (1) the mathematical propositions from which Goldbach's conjecture is derived are true, and (2) Goldbach's conjecture is false. So,

[126] Worley, 20.
[127] Worley, 20.

"every world in which we imagine that Goldbach's conjecture might be false involves a contradiction between the proposition that it is false and these other [mathematical truths]."[128]

The problem, however, is that as limited beings we cannot see that the worlds in which Goldbach's conjecture is false, are necessarily incoherent. That is, the incoherence of these worlds is not at all apparent to us: that is why it seems conceivable that Goldbach's conjecture should be false. And so, even if we accept Chalmers' notion that logical possibility is a matter of conceptual coherence or incoherence, it does not follow that we should be able to detect, simply by examining the concepts available to us, that certain inconsistencies or contradictions obtain.

The implications this has for Chalmers' conceivability argument should now be clear. Given Chalmers' account, it is not enough to consider the proposition that there exist physical duplicates of human beings who lack phenomenal experience to see if that proposition involves any contradiction. In other words, by Chalmers' own light, examining the concepts involved in the proposition that zombies are conceivable is not sufficient to determine whether zombies actually are conceivable. Instead, we must also determine, for every possible world we are considering, whether the truth of this proposition would conflict with any other description of that world. And this may require our knowledge of further concepts that are not yet available to us. As Worley claims, "if it turns out that physical structure actually does entail phenomenal experience then zombie worlds are incoherent, just as worlds in which Goldbach's conjecture is false are incoherent if other mathematical truths mandate the truth of the conjecture."[129] Given that we are limited beings with limited conceptual knowledge, our conceivability judgments are not reliable indictors of what actually is conceivable. Therefore, we cannot be certain that zombies are actually conceivable.

Of course, Chalmers could argue that the situation with regard to Goldbach's conjecture is different than the situation concerning the existence of zombies. Presumably, he would argue that although examining our available concepts *does not* allow us to determine the truth of a complex mathematical puzzle, it *does* allow us to determine that the physical does not necessitate phenomenal conscious experience. This is because, on Chalmers' view, physical

[128] Worley, 20.
[129] Worley, 20.

properties necessitate phenomenal properties only if the phenomenal can be analyzed in structural or functional terms.[130] And this is so because structural and functional properties are the only properties that can be derived from physical properties. Chalmers writes: "physical explanation is well suited to the explanation of *structure* and of *function.* And almost all high-level phenomena that we need to explain ultimately come down to structure or function."[131]

Using an example from Worley (2003), we know that H_2o must be liquid at room temperature precisely because we can give a functional analysis of liquidity. Without that functional analysis, Worley claims, "there would be no way to derive the fact that water must be liquid at room temperature simply from the chemical makeup of water plus the laws of nature."[132] Chalmers, however, argues that a structural or functional analysis of phenomenal properties is not possible, and thus that phenomenal properties cannot be derived from physical properties. He writes:

> But the explanation of consciousness is not just a matter of explaining structure and function. Once we have explained all the physical structures in the vicinity of the brain, and we have explained how all the various brain functions are performed, there is a further sort of explanandum: consciousness itself. Why should all this structure and function give rise to experience? The story about the physical processes does not say.[133]

However, we should be skeptical about this for the same reasons that I earlier argued we should be skeptical about being able to derive conclusions about metaphysical possibility from conceivability (understood as an epistemic notion).[134]

Chalmers' claim that no inference is possible from physical to phenomenal properties must rest on the concepts he has available for thinking about the physical and the phenomenal. For Chalmers to know that no such inference is possible "requires that he know not only that no such inference could be made from our current concepts, but also that these concepts are completely adequate, so that there could be no other description of the situation which might

[130] Chalmers (1996), 107.
[131] Chalmers (1996), 107.
[132] Worley, 22.
[133] Chalmers (1996), 107.
[134] See pages 68-69.

reveal the impossibility of a zombie."[135] However, from the fact that our available concepts do not reveal the possibility that physical properties entail phenomenal properties, it does not follow that no such entailment obtains. Chalmers, therefore, is only in a position to claim that, given our current concepts, it is unclear whether or not physical properties entail phenomenal properties. And this does not warrant his conclusion that no such entailment obtains.

As we have seen, Chalmers' requirements on conceivability are such that we cannot be certain that zombies are conceivable. By Chalmers' own argument, examining the concepts currently involved in the proposition that zombies are conceivable is not sufficient to determine whether zombies actually are conceivable. Instead, as we have seen, we must also determine, for every possible world we are considering, whether the truth of this proposition would conflict with any other description of that world. And this may require our knowledge of further concepts that are not (now) available to us. Yet even if we were to grant Chalmers that zombies are indeed conceivable, it would not be sufficient to show that physicalism is false. In fact, using the same methods that Chalmers employs in his conceivability argument, it is possible to construct an argument from the conceivability of anti-zombies to the truth of physicalism. Let us explore this idea further.

3.4 The Conceivability of Anti-zombies

Chalmers' conceivability argument is intended to dramatize and reinforce his claim that phenomenal properties are irreducible to physical properties. As we have seen, Chalmers argues that the conceptual and thus logical possibility of zombies is sufficient to show that physicalism is false. Using this same line of reasoning, Frankish (2007) argues, however, that the conceptual and thus logical possibility of anti-zombies – purely physical creatures that are nevertheless phenomenally conscious – is sufficient to show that physicalism is true. Both Chalmers' zombie argument and Frankish's anti-zombie argument thus have the same form. And so, if we accept Chalmers' line of reasoning in the zombie argument, we must also accept Frankish's line of reasoning in the anti-zombie argument. This means, therefore, that an argument of this form cannot be used to decide whether we should accept physicalism or reject it.

[135] Worley, 22.

According to Frankish, "an object x [is] a bare physical duplicate of an object y if x is a physical duplicate of y and has no further properties of a non-physical kind."[136] And so, anti-zombies are beings that are bare physical duplicates of us, inhabiting a world which is a bare physical duplicate to ours, but who nevertheless have exactly the same phenomenal conscious experiences as we do. In the anti-zombie world, consciousness is a physical phenomenon logically supervenient on the world's microphysical properties. This supervenience is in virtue of token identities (identities between concrete particulars).[137] And so, what allows a physical property (e.g., consciousness) to supervene on other physical properties (e.g., microphysical properties) is the fact that the physical is composed of different features that are nevertheless themselves physical. Of course, if physicalists are right, then it seems that *we* are anti-zombies. Let us now turn to the argument.

Frankish's anti-zombie argument for physicalism is as follows (the notion of conceivability involved is the same as in the zombie argument):

(1) Anti-zombies are conceivable

(2) If anti-zombies are conceivable, then anti-zombies are metaphysically possible

(3) If anti-zombies are metaphysically possible, then consciousness is physical

Therefore, (4) Consciousness is physical[138]

If we disregard my arguments in sections 3.2 and 3.3 concerning the problems with Chalmers' understanding of conceivability, then premise (1) of the anti-zombie argument is plausible. I am in no way claiming that my previous arguments have no weight or that they can be easily dismissed; rather, the purpose of the anti-zombie argument is achieved by considering it in the same way that *Chalmers* thinks about the zombie argument. And so, given Chalmers' understanding of the notion of conceivability, that is, in terms of conceptual coherence or incoherence, then anti-zombies are conceivable (and thus logically possible). Their description

[136] Frankish, 653.

[137] Here, it is important to distinguish between type identity and token identity. The distinction is an ontological one, between a general sort of thing and its particular concrete instances. Consider, for example, the number of words in the following sentence: "An orange is an orange is an orange." We may count three different words ("an", "orange", and "is") – these words are called types, or we may count eight different words– these words are called tokens.

[138] Frankish, 654.

seems as coherent as that of zombies. My anti-zombie twin, then, is just someone who is physically identical to me, has no non-physical properties, but who (like me) *has* phenomenal conscious experience.

It might be objected that there is an asymmetry between premise (1) of the anti-zombie argument and premise (1) of the zombie argument, in that the former "requires us to embed a totality-clause (i.e., no further properties of a non-physical kind), whereas the [latter] does not."[139] However, it is not clear that this makes premise (1) of the anti-zombie argument less plausible. This is because both premises are similar in the sense that they both require us to conceive of the absence of something – phenomenal properties in one case, non-physical properties in the other.

The case for accepting premise (2) of the anti-zombie argument is the same for accepting premise (2) of the zombie argument; namely, it is an application of the *CP thesis*, the claim that if a situation is conceivable, then it is metaphysically possible. Recall that the jump from conceivability to metaphysical possibility is what physicalists have considered the standard objection to the zombie argument. Here, however, we will accept Chalmers' case against the standard objection – though it is highly controversial (e.g., Brueckner 2001, Block and Stalnaker 1999, and Hill, 1997).

The standard objection to the CP thesis arises from the existence of *a posteriori* necessities. For example, water is necessarily H_2O, but this is an *a posteriori* truth and not *a priori.* That is, it is conceivable (in the epistemic sense) that water is not H_2O, even though water is necessarily H_2O. This shows that the CP thesis is problematic. That is, even though it is conceivable that water is not H_2O, we know a posteriori that water is necessarily H_2O – it is thus logically impossible that water is not H_2O. However, Chalmers argues that conceivability always corresponds to genuine possibility, and reflects the primary intensions of the concepts involved.[140] The primary intension of a concept is the way it specifies what the concept refers to in the actual world, or in any world considered as actual, and "can roughly be thought of as a description of the characteristic features by which we identify the concept's referent."[141] And

[139] Frankish, 654.
[140] For further discussion see Chalmers (1996), 57.
[141] Frankish, 653.

so, the conceivability of water not being H_2O corresponds to the genuine possibility that there is a possible world in which a substance with the identifying features of water (e.g., liquid, odorless, colorless, drinkable, etc.), say glotter, is not H_2O but has some other chemical constituition. On Chalmers' account, we are misled only if we misdescribe this as the possibility that *water* is not H_2O.[142] For Chalmers, explanatory connections must be grounded in a priori entailments from physical facts to high-level facts.[143] Thus, it is the primary intensions of concepts that are relevant in determining whether a given situation is possible. And so, "even if a zombie world is conceivable only in the sense in which it is conceivable that water is not H_2O, that is enough to establish that consciousness cannot be reductively explained."[144] Since Chalmers accepts this line of reasoning for premise (2) of the zombie argument, we should thus also accept it for premise (2) of the anti-zombie argument.

Premise (3) of the anti-zombie argument requires explanation, but like premise (3) of the zombie argument, it is uncontroversial. In the anti-zombie world, consciousness is physical. And so, "the microphysical features of that world (e.g., the laws and distribution of properties) are metaphysically sufficient for consciousness, and any world with the same microphysical features will have the same distribution of phenomenal properties."[145] Notice, by this definition, our world has the same micro-physical features as the anti-zombie world. Thus, the microphysical features of our world are metaphysically sufficient for the existence of phenomenal consciousness. And so, from the conclusion that consciousness is physical, the following is true:

> (5) If anti-zombies are possible, then the microphysical features of our world are metaphysically sufficient for the existence of [phenomenal] consciousness.[146]

It might be objected that all this shows is that we have a physical form of consciousness. But this does not rule out the possibility that we have a non-physical form *as well.* That is, perhaps we possess two otherwise identical sets of phenomenal properties – one that is physical in character, the other non-physical. This objection, however, is easily overcome.

---

[142] For further discussion on Chalmers' position, see Chalmers (1996), 56-71, 98-99, and 131-134.
[143] See for example Chalmers (1996), 98.
[144] Chalmers (1996), 98.
[145] Frankish, 654.
[146] Frankish, 654.

Frankish writes: "even if the suggestion is coherent (and it is not clear that it is), there is absolutely no reason to suppose that it is true, and considerations of simplicity tell overwhelmingly against it."[147] And so, we should accept (5).

This, then, is the anti-zombie argument for physicalism. Although it might seem strange to claim that the mere conceivability of physicalism entails its truth, the general strategy is the same as that employed by Chalmers in the zombie argument. In both the zombie argument and the anti-zombie argument, the CP thesis is used to move from a claim about conceivability (or a psychological claim) to a metaphysical one. In the zombie argument, we are asked to imagine subtracting our phenomenal conscious experiences, while leaving our physical properties intact. From this, we conclude that phenomenal consciousness is non-physical. In the anti-zombie case, on the other hand, we are asked to imagine subtracting our supposed non-physical features while leaving our phenomenal consciousness intact. From this, we conclude that phenomenal consciousness is physical. And so, if the strategy is sound in the one case, then it must also be sound in the other.

In general, the purpose of the anti-zombie argument is to put pressure on the CP thesis. In other words, the intended conclusion is that neither the zombie argument nor the anti-zombie argument is sound. "If both zombies and anti-zombies are conceivable, then [assuming that premise (3) of each argument is accepted] conceivability does not entail [metaphysical] possibility."[148] If we can conceive both of a zombie and an anti-zombie, then it seems that we can conceive both of the existence and non-existence of a being. But, "these conceptions cannot both correspond with genuine possibilities, since a necessarily existing being, exists in all worlds if it exists in any."[149] In other words, a being that necessarily exists is a being that exists in all possible worlds. Just because we can conceive of the non-existence of a necessary being does not mean that it is a genuine possibility that the being does not exist. What is possible, therefore, does not follow from what is conceivable. And so, the CP thesis comes under pressure.

[147] Frankish, 654-655.
[148] Frankish, 655.
[149] Frankish, 655-656.

There are a few different ways that advocates of the zombie argument could reply to the anti-zombie argument. Frankish touches on two possible responses: (1) an appeal to a priori argument for the falsity of physicalism, and (2) the argument that zombies are *ideally conceivable* (i.e., conceivable upon rational reflection) from which it follows, given the CP thesis that anti-zombies are not.[150] It is beyond the scope of this thesis to discuss every objection to the anti-zombie argument. Rather, the purpose of introducing the anti-zombie argument is to demonstrate that even if we accept Chalmers' conceivability argument (i.e., if we ignore my arguments in the first half of this chapter), it is not sufficient to show that physicalism is false – a parallel argument can be made for the truth of physicalism. Nevertheless, there is one possible objection serious enough that I think it worth noting here.

3.5 A Possible Objection to the Anti-zombie Argument[151]

A possible objection to the anti-zombie argument is that it begs the question against interactionist dualism – the view that the mental and physical are fundamentally distinct but interact in both directions (i.e., physical states affect mental states and mental states affect physical states). Frankish writes:

> The idea is that we can conceive of a bare physical duplicate of the actual world only if we assume that the actual world is physically closed under causation. For if some physical events in the actual world are partially determined by facts about a non-physical consciousness, then it is obvious that in a world with the same microphysical laws and the same initial distribution of microphysical properties, events would unfold differently. [152]

At first, this objection seems innocuous, since a similar objection can be made against the zombie argument. That is, the zombie argument requires us to conceive of a physically identical world without consciousness. And so, both the zombie and anti-zombie arguments assume that interactionist dualism is false. However, according to Frankish, "begging the question against interactionist dualism is, arguably, a more serious sin for the [anti-zombie] argument than for [the zombie argument], since the zombie argument can be modified so that it does not beg this question."[153] So, how does the physicalist overcome this objection?

---

[150] For further discussion of these objections and others, see Frankish, 662-665.
[151] This objection was presented to Frankish by Robert Kirk. See Frankish p. 657 footnote 13.
[152] Frankish, 657.
[153] Frankish, 657.

Suppose we have a world that is superficially like ours and which has the same distribution of phenomenal properties, but that is physically closed under causation (any causal gaps being filled by physical substitutes for what are actually non-physical processes); let us call this world w_{pc}.[154] If w_{pc} contains any non-physical properties, they must be epiphenomenal (i.e., have no effect on the physical). Moreover, let us call beings that are physical duplicates of the inhabitants of w_{pc}, but that are not phenomenally conscious, *zombies*$_{pc.}$ Now, the interactionist dualist can run a modified zombie argument to the conclusion that consciousness is non-physical in $w_{pc.}$ For example:

(1) Zombies$_{pc}$ are conceivable

(2) If zombies$_{pc}$ are conceivable, then they are metaphysically possible

Therefore, (3) The microphysical features of w_{pc} are not metaphysically sufficient for phenomenal consciousness[155]

And so, because physicalists will likely say that w_{pc} is the actual world, it follows that physicalism about consciousness is false.

But the same anti-zombie strategy that worked against Chalmers will work against this argument as well. Now, let us call *anti-zombies*$_{pc}$ beings that are bare physical duplicates of the inhabitants of w_{pc}, but that have phenomenal conscious experience. A modified anti-zombie argument can then support the conclusion that physicalism is true in $w_{pc.}$ For example:

(1) Anti-zombies$_{pc}$ are conceivable.

(2) If anti-zombies$_{pc}$ are conceivable, then they are metaphysically possible

Therefore, (3) The microphysical features of w_{pc} are metaphysically sufficient for phenomenal consciousness[156]

However, this conclusion does not entail that physicalism is true. In order to reach that conclusion, an additional premise stating that the actual world is physically closed under causation would be required. Nevertheless, Frankish concludes that "it does cancel out whatever reason the modified version of the zombie argument provides for thinking that physicalism is false, and interactionist dualists must either refute it or cease to employ that

[154] Frankish, 657.
[155] See Frankish, 658.
[156] See Frankish, 658

argument”[157] And so, the anti-zombie argument withstands the objection from interactionist dualism.

[157] Frankish, 658.

Conclusion

As we have seen, the anti-zombie argument neutralizes the zombie argument by showing that the CP thesis is a double-edged sword. That is, the CP thesis can be used to argue from the conceivability of anti-zombies to the truth of physicalism just the same way it can be used to argue against its truth. And so, both the zombie argument and the anti-zombie argument are flawed for the same reason. Moreover, the anti-zombie argument demonstrates that Chalmers' conceivability argument does not overcome the standard objection. However, if we consider the objections made in sections 3.2 and 3.3, then there is a sense in which the anti-zombie argument fails before it even gets off the ground. If we understand the anti-zombie argument in terms of how Chalmers conceives of the zombie argument, then we cannot be sure that anti-zombies are actually conceivable. Examining the concepts involved in the proposition that anti-zombies are conceivable is not enough to determine whether anti-zombies are *actually* conceivable, and thus logically possible.

In chapter three, I have tried to show that there are two ways in which Chalmers' conceivability argument fails. First, given Chalmers' restrictions on conceivability, there is reason to doubt our conceivability judgments. What *seems* conceivable to us is not an adequate indication of what is *actually* conceivable. By Chalmers' own argument, there is reason to doubt the conceivability of zombies. As we have seen, conceptual analysis is not sufficient to determine whether zombies actually are conceivable. Instead, we must also determine, for every possible world we are considering, whether the truth of this proposition would conflict with any other description of that world. And this may require our knowledge of further concepts that are not available to us.

Second, as the anti-zombie argument demonstrates, Chalmers' conceivability argument does not overcome the standard objection that conceivability cannot serve as a guide to, or evidence for, metaphysical possibility. The CP thesis can also be used to argue for the truth of physicalism. And so, the anti-zombie argument can be defeated only by rendering the zombie argument itself redundant. Dualists, therefore, should not be advocates of the zombie argument. Frankish captures the essence of the anti-zombie argument when he writes: "when zombies and anti-zombies meet, they annihilate each other, and in so doing, reveal that

considerations of conceivability have little role to play in debates about the nature of consciousness."[158]

Chalmers is a leading figure in the quest to understand the nature of consciousness. Although his work has been extremely influential, his dualist position is not enough for a complete theory of consciousness. As this thesis shows, Chalmers has more work ahead of him if he is to adequately argue that phenomenal consciousness is non-reducible. To be fair, however, reductionists also have an equally long road ahead of them. It is not enough to simply show that phenomenal states *correspond* to physical states; rather, if physicalism is to be validated, it must be shown that phenomenal properties are reducible to physical properties.

In this thesis I have tried to show why Chalmers' argument against a physicalist theory of consciousness fails. First, his argument that consciousness fails to logically supervene on the physical rests on an ambiguous interpretation of logical supervenience. On one hand, Chalmers understands logical supervenience as an ontological relation of determination, on the other, as an epistemological relation whereby knowledge of the A-facts entails knowledge of the B-facts. Chalmers does not differentiate between these two interpretations, and instead treats them as equivalent. He is thus guilty of problematically equivocating between the two. And so, unless Chalmers remedies this, his argument should not be taken seriously.

Second, Chalmers' conceivability argument, though seductive (given the vast amount of literature on zombies) is nevertheless flawed. Given his restrictions on conceivability, it is not at all apparent that zombies are conceivable. Moreover, his argument does not seem to overcome the standard objection that conceivability cannot serve as a guide to, or evidence for, metaphysical possibility. As the anti-zombie argument demonstrates, it is possible to formulate an argument using Chalmers' CP thesis but which argues for the truth of physicalism. Given that physicalists are able to construct this argument, it should be clear to Chalmers that we cannot determine the nature of consciousness through an exercise of the imagination alone.

For the time being, dualists and reductionists are gridlocked – each side presents compelling arguments for their position, but clearly neither is able to fully grasp or explain the nature of consciousness. For now, the problem of consciousness lies uneasily at the border of

[158] Frankish, 666.

philosophy and science. Contra Chalmers, I do not think that the study of phenomenal consciousness evades empirical study. In fact, I think both philosophical investigation and empirical study is necessary if we want to learn the truth about consciousness.

Bibliography

Armstrong, David. *A Materialist Theory of the Mind.* London: Routledge and Kegan Paul, 1968.

Balog, Katalin. "In Defense of the Phenomenal Concept Strategy." *Philosophy and Phenomenological Research*. Forthcoming, 2010.

_____. "Zombies, Illuminati, and Metaphysical Gridlock." Unpublished MS. 13 Apr. 2010. Available from: <http://pantheon.yale.edu/~kb237/Web%20publications/ZRedux.pdf>.

Blackburn, Simon. "Moral Realism." *Morality and Moral Reasoning.* Ed. J. Casey. London: Methuen, 1971.

_____. "Supervenience Revisited." In *Exercises in Analysis: Essays by Students of Casimir Lewy.* Ed. I. Hacking. Cambridge: Cambridge University Press, 1985.

Block, Ned and R. Stalnaker. "Conceptual Analysis, Dualism, and the Explanatory Gap." *Philosophical Review*. 108 (1999): 1-46.

Broad, C.D. *The Mind and its Place in Nature.* New York: The Humanities Press Inc., 1925.

Brown, Richard. "Deprioritizing the A Priori Arguments Against Physicalism." *Journal of Consciousness Studies.* 17 (2010): 1-23.

Brueckner, Anthony. "Chalmers's Conceivability Argument for Dualism." *Analysis*. 61.3 (2001): 187-193.

Chalmers, David. "Facing Up to the Problem of Consciousness." *Journal of Consciousness Studies*. 2.3 (1995): 200-219.

_____. "The Puzzle of Conscious Experience." *Scientific American*. (1995): 90-100.

_____. *The Conscious Mind: In Search of a Fundamental Theory.* New York: Oxford University Press, 1996.

_____. "*Response to Searle."* 1997. 13 Dec. 2009 <http://consc.net/book/searle-response.html>.

_____. "Materialism and the Metaphysics of Modality." *Philosophy and Phenomenological Research.* 59.2 (1999): 473-496.

_____. "Does Conceivability Entail Possibility?" In *Conceivability and Possibility.* Eds. T. Gendler and J. Hawthorne. Oxford: Oxford University Press, 2002.

_____. “Consciousness and Its Place in Nature.” In *Philosophy of Mind: Classical and Contemporary Readings.* Ed. David Chalmers. Oxford: Oxford University Press, 2002. 247-271.

_____. “How Can we Construct a Science of Consciousness?” *The Cognitive Neuroscience III.* Ed. M. Gazzaniga. Cambridge: MIT Press, 2004.

_____. “Phenomenal Concepts and the Knowledge Argument.” In *There’s Something About Mary: Essays on Phenomenal Consciousness and Frank Jackson’s Knowledge Argument.* Eds. P. Ludlow, Y. Nagasawa, and D. Stoljar. Cambridge: MIT Press, 2004.

_____. “Strong and Weak Emergence.” In *The Re-Emergence of Emergence*. Eds. P. Clayton and P. Davies. Oxford: Oxford University Press, 2006.

_____. “Phenomenal Concepts and the Explanatory Gap.” In *Phenomenal Concepts and Phenomenal Knowledge: New Essays on Consciousness and Physicalism.* Eds. T. Alter and S. Walter. Oxford: Oxford University Press, 2006.

_____.“The Two-Dimensional Argument Against Materialism.” In *Oxford Handbook of the Philosophy of Mind. Ed*. B. McLaughlin. Oxford: Oxford University Press, 2009.

Churchland, Paul M. “The Rediscovery of Light.” *Journal of Philosophy*. 93 (1996): 211-228.

Churchland, Patricia S. “The Hornswoggle Problem.” *Journal of Consciousness Studies*. 3 (1996): 402-408.

Clark, Thomas. “Function and Phenomenology: Closing the Explanatory Gap.” *Journal of Consciousness Studies*. 2 (1995): 241-254.

Dennett, Daniel. “Facing Backwards on the Problem of Consciousness.” *Journal of Consciousness Studies*. 3 (1996): 4-6.

Descartes, René. In *The Philosophical Writings of Descartes*. Trans. John Cottingham, Robert Stoothoff, & Dugald Murdoch. 2 vols. Cambridge: Cambridge University Press, 1984.

Descartes, René. “Passions of the Soul.” In *Philosophy of Mind: Classical and Contemporary Readings.* Ed. David Chalmers. Oxford: Oxford University Press, 2002. 21-23.

Feigl, Herbert. “The Cognitive Role of Acquaintance.” In *Minnesota Studies in the Philosophy of Science II: Concepts, Theories, and the Mind-Body Problem.* Eds. Herbert Feigl, Michael Scriven, and Grover Maxwell. Minneapolis: University of Minnesota Press, 1958.

Frankish, Keith. “The Anti-zombie Argument.” *The Philosophical Quarterly*. 57.229 (2007): 650 – 666.

Hardcastle, Valerie. “The Why of Consciousness: A Non-issue for Materialists.” *Journal of Consciousness Studies*. 3 (1996): 7-13.

Hare, Richard M. “Supervenience.” Proceedings of the Aristotelian Society. 58 (1984): 1-16.

Hill, Chris. “Imaginability, Conceivability, Possibility, and the Mind-Body Problem.” *Philosophical Studies*. 87 (1997): 61-85.

_____ and Brian McLaughlin. “There are Fewer Things in Reality Than are Dreamt of in Chalmers’s Philosophy.” *Philosophy and Phenomenological Research*. 59 (1999): 446-454.

Hodgson, David. “The Easy Problems Ain’t So Easy.” *Journal of Consciousness Studies*.3 (1996): 69-75.

Jackson, Frank. “What Mary Didn’t Know.” *Journal of Philosophy*. 83 (1986): 291-295.

_____. *From Metaphysics to Ethics*. New York: Oxford University Press, 1998.

Kim, Jaegwon. “Concepts of Supervenience.” *Philosophy and Phenomenological Research*. 45.2 (1984): 153-176.

Kirk, Robert. *Zombies and Consciousness.* New York: Oxford University Press, 2005.

Kripke, Saul. “Naming and Necessity.” In *Semantics of Natural Language.* Eds. D. Davidson and G. Harman. Dordrecht: Reidel, 1972. [Reprinted as Kripke 1980].

_____. *Naming and Necessity.* Cambridge: Harvard University Press, 1980.

Kusch, Martin. *Psychological Knowledge: A Social History and Philosophy.* New York: Routledge, 1999.

Laplace, Pierre-Simon. *A Philosophical Essay on Probabilities.* Trans. F.W. Truscott and F.L. Emory. New York: Cosimo, Inc., 2007.

Latham, Noa. “Chalmers on the Addition of Consciousness to the Physical World.” *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. 98.1 (2000): 71-97.

Levine, Joseph. *Purple Haze: The Puzzle of Consciousness.* New York: Oxford University Press, 2001.

_____. “Conceivability and the Metaphysics of Mind.” *Noûs*. 32 (1998): 449-480.

Lewis, David. “An Argument for the Identity Theory.” *Journal of Philosophy*. 63 (1968): 17-25.

Loar, Brian. “David Chalmers’s *The Conscious Mind*.” *Philosophy and Phenomenological Research*. 59 (1999): 465-472.

Moore, G.E. *Philosophical Studies*. London: Routledge and Kegan Paul, 1922.

Nagel, Thomas. “What Is It Like to Be a Bat?” *Philosophical Review*. 83.4 (1974): 435-450.

Passer, Michael et al. *Psychology: Frontiers and Applications.* 2nd ed. Toronto: McGraw-Hill Ryerson, 2005.

Putnam, Hilary. “Minds and Machines.” In *Dimensions of Mind*. Ed. Sidney Hook. New York: New York University Press, 1960.

Robinson, Howard. “The Anti-materialist Strategy and the Knowledge Argument.” In *Objections to Physicalism*. Oxford: Oxford University Press, 1993.

Rowlands, Mark. *The Nature of Consciousness.* Cambridge: Cambridge University Press, 2001.

Ryle, Gilbert. *The Concept of Mind*. London: Hutchinson, 1949.

_____ . “Descartes’ Myth.” In *Philosophy of Mind: Classical and Contemporary Readings*. Ed. David Chalmers. New York: Oxford University Press, 2002. 32-38.

Searle, John. *The Rediscovery of the Mind.* Cambridge: MIT Press, 1992.

Shoemaker, Sydney. “Functionalism and Qualia.” *Philosophical Studies*. 27 (1975): 291-315.

_____ . “On David Chalmers’ ‘Conscious Mind.’” *Philosophy and Phenomenological Research*. 59 (1999): 439-444.

Van Cleve, James. “Conceivability and the Cartesian Argument for Dualism.” *Pacific Philosophical Quarterly*. 64 (1983): 35-45.

Warner, Richard. “Facing Ourselves: Incorrigibility and the Mind-Body Problem.” *Journal of Consciousness Studies*. 3 (1996): 217-230.

Weisstein, Eric. W.”Goldbach Conjecture.” From *Math World: A Wolfram Web Resource*. 2010. 10 Apr. 2010 <http://mathworld.wolfram.com/GoldbachConjecture.html>.

White, Stephen. “Curse of the Qualia.” *Synthese*. 68 (1986): 333-368.

Worley, Sara. “Conceivability, Possibility, and Physicalism.” *Analysis*. 63.1 (2003): 15-23.

Yablo, Stephen. “Concepts and Consciousness.” *Philosophy and Phenomenological Research*. 59.2 (1999): 455-463.

_____. “Is Conceivability a Guide to Possibility?” *Philosophy and Phenomenological Research*. 53 (1993): 1-42.